

STIC-Biotech/ChemLib

72221

From: Slobodyansky, Elizabeth
Sent: Wednesday, July 31, 2002 11:44 AM
To: STIC-Biotech/ChemLib
Subject: 09/900,038

Please search for case 09/900,038:

SEQ ID NOs: 1 and 2 against commercial and ^{ES} ~~interference~~ databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652
CM1 10D11
703-306-3222

mail box 10C01

CRFE

Searcher: ^{Point of Contact} ~~P Sheppard~~
Phone: ^{phone number: (703) 308-4499}
Location: _____
Date Picked Up: _____
Date Completed: 8/6/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 10:43:39 ; Search time 1790.74 Seconds
(without alignments)
10973.127 Million cell updates/sec

Title: US-09-900-038A-2

Perfect score: 939

Sequence: 1 atgaattatgattcattat.....gaggagagaacaaagtgc 939

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_cm.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

AB050723

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

AB050723

Streptococcus agalactiae cpsIBD, cpsIBF, cpsIBG, cpsIBH, cpsIBI,

cpsIBJ, cpsIBK, cpsIBL, neuB, neuC genes, complete cds.

AB050723.1 GI:12697593

Streptococcus agalactiae DNA.

Streptococcus agalactiae

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

1 (sites)

Miyake,K., Watanabe,M. and Iijima,S.

CpsJ of Streptococcus agalactiae type Ib shows

beta-1,3-galactosyltransferase activity

Unpublished

2 (bases 1 to 9987)

Miyake,K., Watanabe,M. and Iijima,S.

Direct Submission

Submitted (01-NOV-2000) Katsuhide Miyake, Nagoya University, Grad.

Sch. of Engineering, Dept. of Biotechnology; Chikusa-ku, Furo-cho,

Nagoya, Aichi 464-8603, Japan

AB050723 Streptoco
AF337958 Streptoco
U67549 Methanococ
AF285774 Bacteroid
AE007717 Clostridi
AF279134 Dictyoste
AL031746 Bacillus
AF001519 Plasmodi
AC092304 Homo sapi
AF402315 Shigella
Y07786 V.cholerae
AF461121 Escherich
AB012956 Vibrio ch
AC106820 Homo sapi
AX344555 Sequence
AC104027 Homo sapi
AC012492 Homo sapi
AC005504 Plasmodi
AC004157 Plasmodi
AC008079 Homo sapi
U32832 Haemophilus
AF400048 Campyloba
AF361371 Escherich
AF343914 Campyloba
AC096750 Homo sapi
AC009277 Homo sapi
AC011718 Homo sapi
AL359713 Human DNA
AC006970 Homo sapi
AC093790 Homo sapi
AC027753 Homo sapi
AC066608 Homo sapi
AC068139 Homo sapi
AC104814 Homo sapi
AC005504 Plasmodi
AC004157 Plasmodi
AX347057 Sequence
AC020549 Homo sapi
AC034148 Homo sapi
U11584 Drosophila
AE005429 Escherich
U37541 Drosophila
AB008676 Escherich
AC079621 Homo sapi
AP002559 Escherich

(E-mail:miyake@proc.nubio.nagoya-u.ac.jp, Tel:81-52-789-4278,
Fax:81-52-789-3221)

FEATURES	source	Location/Qualifiers	
		1..9987	
		/organism="Streptococcus agalactiae"	RBS
		/db_xref="taxon:1311"	gene
RBS		425..430	
gene		442..1791	CDS
		/gene="cpsIbD"	
CDS		442..1791	
		/gene="cpsIbD"	
		/function="glucosyltransferase"	
		/codon_start=1	
		/transl_table=11	
		/evidence=not_experimental	
		/product="CpsIbD"	
		/protein_id="BAB21597.1"	
		/db_xref="GI:12697594"	
		/translation="MIQTVVYFVSASLTLTITPNFKSNKOLLFVLLIHIVFYLSDF YRDMRGYLEEFKMLVKYSFYIPISSLSFFISKNSTTTTSLSPFTFIAMNSLILYL LNSFLYKRYKYSYAKFSRDTKVLITNKSLSKMTFRNKYDHNHYIAVCLIDSSKDCY DLKHNLSRAIINKDALTSBTLCTVDAQFINIPIELFGYQIQDIINDIEAMGVIVNVN VEALSFNIGEKRIQTFEGYSVITYSMKEYKSHLIAKREFLDIMGAILLIGCIVAI FLVPIRKDGPAPFSQNRVGRNGRIFRFRFSRMKVDAEQKKOLLVHNQMTLMFK LDDPRIITKIGKFIKRTSIDELPQFINVLKGDMSLVGTRPPTVDYERYNSTQKRRLS FKPGITGLWQISGRNNITDFDEIVKLDVQYINWSIWSIDIKIILITLKVLLGTGAK"	
RBS		1803..1810	
gene		1815..2264	
		/gene="cpsIbF"	
CDS		1815..2264	
		/gene="cpsIbF"	
		/codon_start=1	
		/transl_table=11	
		/evidence=not_experimental	
		/product="CpsIbF"	
		/protein_id="BAB21598.1"	
		/db_xref="GI:12697595"	
		/translation="MKICLVSSGGHLAHLNLLKPIWEKEDRFWVTFDKEDARSILRE ETVYHCFPTNRNVKNVKNLTIAFKVLKRKRPDVISSGAAVAPFFVIGKLFCKT VYIEVFDRIDPTLGLVYPVTDKFIQVWEEMKKVYPKAINLGIF"	
RBS		2249..2256	
gene		2264..2746	
		/gene="cpsIbG"	
CDS		2264..2746	
		/gene="cpsIbG"	
		/function="beta-1,4-galactosyltransferase"	
		/codon_start=1	
		/transl_table=11	
		/evidence=not_experimental	
		/product="CpsIbG"	
		/protein_id="BAB21599.1"	
		/db_xref="GI:12697596"	
		/translation="MIFVTVGTHEQFNRLIKEVDRLKGTGAIDQEVFIQTGYSDFEP QNCQWSKFLSDYMDMSYKAEIVITHGGPATFMNAVSKGRKTIIVVPRQEPQGEHVN HOVDLKLFLKYELDYILNTSELENIIEKNISTSKVISQNNDFCSFKNLSKLFE"	
RBS		2806..2812	
gene		2842..3981	
		/gene="cpsIbH"	
CDS		2842..3981	
		/gene="cpsIbH"	
		/function="putative CP polymerase"	
		/codon_start=1	
		/transl_table=11	
		/evidence=not_experimental	
		/product="CpsIbH"	
		/protein_id="BAB21600.1"	
		/db_xref="GI:12697597"	
		/translation="MQLLLALIVLLICSSYNEKMKFLNMAEIFFIVFYMYLVLSIV LNSLFRSPEFHVIAFNSLAVGVVSLFYHYVKNTELKLLKSLFNALILFCIG FLYYAIYFDVENVSLFGRNLIGSDWINGMHTORAMAFEYSNLIPLITIINIYII YIKQYISSGMILGALLSTIILPIGGSGRAGIIVVLLQVLIILLNTIIVIKROTIRFEL	
RBS		3964..3970	
gene		3981..4955	
		/gene="cpsIbI"	
CDS		3981..4955	
		/gene="cpsIbI"	
		/function="beta-1,3-N-acetylglucosaminyltransferase"	
		/codon_start=1	
		/transl_table=11	
		/evidence=not_experimental	
		/product="CpsIbI"	
		/protein_id="BAB21601.1"	
		/db_xref="GI:12697598"	
		/translation="MEKQLIVSIVIPYNSEAYLKECVOSVLOQTHSLIEVLINDGS TDNSEICDNLSQKDDRILVPHKKNGVSSARNLGLDKSTGEFITFVDSDDFAPNII EIMLNLTEDADIAEVDIFDINERDYRKKRRNFYKVFKNNSLKFEFSGNRVENIV CTKLYKSLIGNLREFDENLKIGEDLLFNCKILCQEHCIWVDTTSSLYTVRIYKTSAMN QEFNSLDQFITFENEISSIVPAKLANVEAKFLREKVKCLRMKMFELGNSIDSKIKIQ REIFFKDKVLFPFYKAVYLSLGLLSIYLMKCSPLYIKLYDRFQKQ"	
RBS		4995..5001	
gene		5008..5949	
		/gene="cpsIbJ"	
CDS		5008..5949	
		/gene="cpsIbJ"	
		/function="beta-1,3-galactosyltransferase"	
		/codon_start=1	
		/transl_table=11	
		/evidence=experimental	
		/product="CpsIbJ"	
		/protein_id="BAB21602.1"	
		/db_xref="GI:12697599"	
		/translation="MNYSIIMSVYNEPLNVRDVSILNQTLTDFEFTIIVDNPISRG DLKQFLTYSVVNDRIKILLNEENIGLASSLNKAVKISGEYIFRMDADDISYPSRFD KOIRMEANSLDFSATLIELIDQKGNLVYKQRESNKIYIINDIRMLNRSILAHPTW CVKVKVQKMGYRDLPVEDYDFAIRGALADFKIGLLNKVLLQVLRNENGISQTNKF KQYIYSAILODFYKESYIDITKITNYFOEYVIKKRYTOQLSKYFELKSTPSITIRK LYICLVLKFSPLVRRLLINDINILVLKLGGEKQSD"	
RBS		5928..5935	
gene		5942..6880	
		/gene="cpsIbK"	
CDS		5942..6880	
		/gene="cpsIbK"	
		/function="putative sialyltransferase"	
		/codon_start=1	
		/transl_table=11	
		/evidence=not_experimental	
		/product="CpsIbK"	
		/protein_id="BAB21603.1"	
		/db_xref="GI:12697600"	
		/translation="MTNRKIYVCHTLYHLLICLYKEEYISNLEIILSSSIDVDNLEK KLKSTINIHILESSGESELLVKLDAGLSYKFDNSCNCFIFNDATPIGRLLIKRGI YLNLEDGLNCFTYIFSQKLMKYVKKYILHKIQPHGFSRYCLGLIEVNSLVNPKDP RYKKEIYPRKELFDNVEYQKEMAINLFGAVRVSITKSPSVLVLQPLSIDKEFMSYN NKIETSEQNFYKSVINVEYINKGVNVLKVPDRVDYDYSKLPVELLPSNVPMIEEL MLTGREGGITHSSTALDFTCVDKKITVLVDLKDIK"	
RBS		6865..6870	
gene		6877..8277	
		/gene="cpsIbL"	
CDS		6877..8277	
		/gene="cpsIbL"	
		/function="putative repeat unit transporter"	
		/codon_start=1	
		/transl_table=11	
		/evidence=not_experimental	
		/product="CpsIbL"	
		/protein_id="BAB21604.1"	
		/db_xref="GI:12697601"	
		/translation="MNTKKLQSGFIYITIGNLLVOGLAFITLPIYTRVISTEYVGQYS	

```
LYVAMNMIIMFLICLOTSGSSARVKYGEERKSYSGSAFSGVNIWFLVLLIAFLFR
SLFAPLWDSFESIFLLMVCOSYANVYVTFQCVYFIOORSLANLILSLANAVSSVALS
LFLEFHWSDDFLRLVGLVFVPTLIITGIVAFAYIYHKSFSYKPYFPIVTVSPLIF
HLHGOLLNOLDRLMARLYSTKEVAMYSFGYSLGHMIIQIYVLANSINMANWIPWEDARK
EKLLQSTYSIRLYLGLVFLUTGLTFVFPBQAQIMGGDQYSSSQFISLLIYSFLVF
LYTFPNNIOFYFANTWNPITGLLAAGVNNLNLVLIPIHYAAGAAATIIYSIALLI
FHHVSKVKYHYSDVSVRQYIILSGIVFSYAMLNMFNDINIVRWSLGIILIVYSIV
FORVILDLSKKRRR"
RBS 8262..8269
/gene="cpsIBL"
gene 8277..9302
/gene="neuB"
CDS 8277..9302
/function="sialic acid synthetase"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="NeuB"
/protein_id="BAB21605.1"
/db_xref="GI:12697602"
/translators="MYIIAEICNHNGDINLAKKVDVAVSCGVDAVKFQTPKAEL
ISFAPAEYOKATTTGADSOLEMKRLEUSFEIYLEMRYAISKGVETFTSPDEES
LEFLISTDMPDIYKIPGEITNLPYLEIKQKRVILSTGMAYMEEIHOAVNLRQNG
TTDISILHCTEYTPYPSLNLNVIHTLKDFKDLTIGYSDHSTGSEVPFAAAMGAE
VIBKHTLDTNMEVDPDHKASATPDILAAVLKGFALLNQALGRFEKIPDPVEEKNKIVA
RKSVALKPIKKGDIYSIENITVKRPGNGISPMNWDILQEAQDDFEDEVIRDSF
ENQLPEL"
RBS 9367..9374
gene 9379..9987
/gene="neuC"
CDS 9379..9987
/gene="neuC"

Query Match 100.0%; Score 939; DB 1; Length 9987;
Best Local Similarity 100.0%; Pred. NO. 2.5e-134;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaattatagatcattatgcgtatataatagagccttaaatattatgtgagagattca 60
Db 5008 ATGAATTATAGTATCATATGTCGGTATATATGATGAGCCTTTAAATATATGAGAGATTCA 5067

Qy 61 gtagaattatattaaatcaaacgcttactgattttgagttcataaattgtcattgataat 120
Db 5068 GTAGAATCTATATTAATCAAAACGCTTACTGATTTTGAGTTTCATAATTGTCATTGATAAT 5127

Qy 121 ccaagtagaggtgatttaaacgaattcttcaacagaattcagttgtagataatagaata 180
Db 5128 CCAAGTAGAGGTGATTTAAAGCAATTTCTTAACAGAATATTTCAGTTGTAGATAATAGAATA 5187

Qy 181 aaastcttgcttaagaataatttggtttagcatcaagatttgaacaaagcggtgaaa 240
Db 5188 AAAATCTTGCTTAATGAAGAAAATATTTGGTTTACATCAAACTTTGAACAAACCGGTGANA 5247

Qy 241 atttcaaggagagaataatttttagaattgagatgctgatgatatttcataatccaagttaga 300
Db 5248 ATTTCTAAGGGAGAATATATTTTAGAATGGATGCTGATGATATTTTCATATCCAAGTAGA 5307

Qy 301 tttgataagcaaaatcgtttatttagagagaaattcattggattctcagcaactctaata 360
Db 5308 TTTTGATAAGCAAAATTCGTTTTATGGAGAAAATTTCAATGGATTTCTTCAGCAACTTAATA 5367

Qy 361 gaattgatagcacaaaggaatttagtatataaacaacagagaagaataataataatatac 420
Db 5368 GAATTGATAGACCAAAAGGAATTTTAGTATATAAACACGAGAAAGTAATAAATATATAC 5427

Qy 421 ttaactaatgatatacagaagatgatttgaatagatctatctacttgcaccaccaactcgg 480
Db 5428 TTAACATGATATACGAAGATGTTATTGAATAGATCTATACTTGCACCACCAACGGTGG 5487

Qy 481 tgcgtaaaaaagaagtttttcgataaagttaatgggatataagagatttagtacctgttgaa 540
Db 5488 TGCGTAAAAAAGAAGTTTTCGATAAGTTAATGGGATATAGAGATTTAGTACCTGTTGAA 5547
```

```
Qy 541 gattatgattttcaataagagagctctgcgtatttcaaaatcgcttactcaataaaa 600
Db 5548 GATTATGATTTTGCATTAAGAGGAGCTCTGGCTGATTTCAAAATCGGCTTACTCAATAAA 5607

Qy 601 gtactttcacagtagattataaacgagagatggaatcacacaaacccaataaagtttaagcaa 660
Db 5608 GTACTTTTACAGTAGATAGATTAAACGAGAAATGGAATATACAAAACCAATAAAGTTTAAGCAA 5667

Qy 661 tatatttactcagctatttttcaagagattttataaagaaaaaatcttatatgatatacaca 720
Db 5668 TATATTTTACTCAGCTATTTTACAAGATTTTATAAAGAAAAATCTTATATTGATATCACA 5727

Qy 721 aaaaattactaattactttccaagagatgatgataaagaacgctatactacgacgaagctc 780
Db 5728 AAAATTACTAATTACTTTCAAGAGTATGTGATAAAGAANAACGCTATACACTCAGCAAGGCTC 5787

Qy 781 tctaaatattttgagctaaaaatctacccctagtagtattactattagaaaaactatataattgt 840
Db 5788 TCTAAATATTTTGAGCTAAAATCTACCCCTAGTATTACTATTAGAAAACTATATATTGT 5847

Qy 841 ttataatttactatttaagctcccttggttaggaggttattataaataaataatgataattatt 900
Db 5848 TTATATTTTATATCTTTAAAGTCTCCCTTGCTTAGGAGGTTATTAAATAAATGATATTATATT 5907

Qy 901 ttagtactgaaattggtttggagagagaaacaaaagtgcac 939
Db 5908 TTAGTACTGAAATTTGTTGGAGGAGAGAAACAAGTGAC 5946

RESULT 2
AF337958 16448 bp DNA linear BCT 27-FEB-2001
LOCUS Streptococcus agalactiae strain NT6 type VI capsular polysaccharide
DEFINITION synthesis operon, complete sequence.
ACCESSION AF337958
VERSION AF337958.1 GI:13022158
KEYWORDS
SOURCE Streptococcus agalactiae.
ORGANISM Streptococcus agalactiae.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 16448)
AUTHORS McKinnon,K., Chaffin,D.O. and Rubens,C.E.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) Infectious Disease, Immunology and
Rheumatology, Children's Hospital and Regional Medical Center, 4800
Sand Point Way NE, Seattle, WA 98105, USA
FEATURES
source
1..16448
/organism="Streptococcus agalactiae"
/strain="NT6"
/serotype="VI"
/db_xref="taxon:1311"
895..16444
/gene="type VI capsular polysaccharide synthesis operon"
895..2425
/gene="cpsA"
895..900
/gene="cpsA"
919..924
/gene="cpsA"
956..961
/gene="cpsA"
968..2425
/gene="cpsA"
/note="similar to LytR"
/codon_start=1
/transl_table=11
/product="CpsA"
/protein_id="AAK11658.1"
/db_xref="GI:13022158"
/translation="MSNHSRRHQRKHSHTPLRVINLFLVIFILLVSVSLFLMYRHHF
```

```
LAFRHLNVIYGVWIIILASLFLCICKKAKRIFTTIIILVLSISFVATVTLYGFKSTIDL
TNNLKTSYSEIENSVVVPKDSKITNIEAVSKLAAPVKNDTSNITDLIEHKSEKGI
SITPKOTSDYDAYNRIKNGSQAMVNNAYVSLIELSTPDFKSOIKTIYTIYKIKKKI
NRKNTNHEGVFNIIYISGIDTFGSISTVSRSDVNIIMVTNTNTHKVLTTTPRDAYVK
IPDGGNDYDKLTHAGLXGVETSMKTLNLDYDVARINFSSFLKLIDLGGTVV
YNDQAFTHKGNDFPVGQVTLNSEQALGFVRERYSLOGGNDGRNQEKVIAALINK
LASSQSVTKLNSITTSQLOTSVOTNMTINDINLNQISTQORFTVESQALTGHGSGT
ELPSYMPGAQOLYMMSIDOSSLSNAKSIIKNTMEP"
2415. .2421
/ gene="cpsB"
2415. .3162
/ gene="cpsB"
2431. .3162
/ gene="cpsB"
/ codon_start=1
/ transl_table=11
/ product="tyrosine phosphatase CpsB"
/ protein_id="AAK11659.1"
/ db_xref="GI:13022160"
/ translation="MIDIHSHIVDFVDGPKTLEESLSLIEESYRQGVRIIVSTSHRR
KGMFTPEDDIIFKNFSIVKHEAKREFHLQILYGGELYTSDMLEKCLKLQIPTLNNT
KFALTEFSMOTSKDIIHAKLSNVMLGITPVVAHLERYNALENOKERYKEIINMGCYT
QINSSHILKOLFNKDKHKKRARYFLEENLVHVFASDMHNLDRVPPFLAEAYKIC
RDFGKERANQLFIENQAQSLKNHYI"
3161. .3863
/ gene="cpsC"
3161. .3167
/ gene="cpsC"
3171. .3863
/ gene="cpsC"
/ codon_start=1
/ transl_table=11
/ product="putative chain length regulator CpsC"
/ protein_id="AAK11660.1"
/ db_xref="GI:13022161"
/ translation="MNKIANTEVEINFIENLLKMKKKELTFFVAIAFATAGLFVSLF
IVTPQYTSSTRIYVNPNTPNNSITQAQDQAGSFLANDYKEIITSDVLEKVISSEKL
NYPSSQLQKIDTISLKDRVISISVEDANPKMSOKLANSVREAAVSKIKAVTQVEDI
TTLEKGNLPKAPSSPNIKKNVLIGFVAGLSTIVLVMIGLIDLRVNTDEETEKVLGL
TSLGIVPDLNKL"
3863. .4563
/ gene="cpsD"
3863. .3868
/ gene="cpsD"
3874. .4563
/ gene="cpsD"
/ codon_start=1
/ transl_table=11
/ product="capsular polysaccharide chain length
regulator/exporter CpsD"
/ protein_id="AAK11661.1"
/ db_xref="GI:13022162"
/ translation="WTRLEIVDSKLROAKKTEEFNAIRNTIOFSGKKNKILAITSVR
EGEKSSTTSLSLSLAQAGPKTLIIDADTRNSVMSGTFKATGIKTLITNLSGNADL
GDIICTETPRLMVVPVSGVPPNPNTALLQNAIFNKMIEAIRNIFDIILIDPPGLV
DAAIIASACDGFVLVTQAGRIKRNVEKAKEQMEQSGSKFLGIILNKNVESVATYGDY
GNYGRDRKK"
4577. .4582
/ gene="cpsE"
4577. .5964
/ gene="cpsE"
4615. .5964
/ gene="cpsE"
/ codon_start=1
/ transl_table=11
/ product="glucose-1-phosphate transferase CpsE"
/ protein_id="AAK11662.1"
/ db_xref="GI:13022163"
/ translation="MIQTWVWVSVSLTTLITPNFKSNKDLFLVLLIHYIVFLSDF
YRDFWRCGLBEFKMVLKYSFYIFISSLFFIFKNSTFTRLSPTFIAMNSILLYL
LNSFLKYRYKYSYAKFSRDTKVLITNKDSLKMTFRNKYDHNITAVCILDSSEKDCY
DLKHSNRIINKDALTSELTCLTVQDAFINIPIELFGKYQIODIIDIEMGVYVNW
VEALSFDNIGEKRIQTPEGYSVITYSMRFYKYSHLIARFRUIDIMGAILGLLTCIGVAI
```

```
FLVPOIRKDDGGAIPFSQNRVGRNGRIFRFYKFRSMRVDAEQIKKOLLVHNQMTGLMFK
LEDDPRITKIKGFIKRTKSIDELPQFYNVLKGDMSLVGTRPPTVDEYKYNSTQKRRLS
FKPGITGTGWSIGRNNITDFDEIVKLDVQYINEMSIWSDIKILITLTKVLLGTAK"
5975. .6437
/ gene="cpsF"
5975. .5981
/ gene="cpsF"
5988. .6437
/ gene="cpsF"
/ codon_start=1
/ transl_table=11
/ product="beta-1,4-galactosyltransferase enhancer CpsF"
/ protein_id="AAK11663.1"
/ db_xref="GI:13022164"
/ translation="MKICLVSSGGHLAHLNLLKPIWEKEDRFWTFOKEDARSILRE
EIVYHCFPTNRNVKNVLTILAFKVLKRKRPDVIISGGAAVAPFFYIGLFGCKT
VYIEVDFRIDKPTLTGKLVYPVTDKFIQVWEEMKKIYPKAINLGGIF"
6422. .6428
/ gene="cpsG"
6422. .6910
/ gene="cpsG"
6437. .6910
/ gene="cpsG"
/ codon_start=1
/ transl_table=11
/ product="beta-1,4-galactosyltransferase CpsG"
/ protein_id="AAK11664.1"
/ db_xref="GI:13022165"
/ translation="MIFVTVGTGHEQQFNRLIKEVDRLKGTGAIDQEVFIOTQYSDFEP
QNCQWSKFLSYDDMSYMKAEIVITHGGPATFMSVISLGKLPVVPVPRKKQGEHIND
HQIQLFNSIAHLIPLAWIEDVDGLAEALKRNATEKYOQGNNDMFCKLEKIIGET"
6898. .8064
/ gene="cpsH"
6898. .6903
/ gene="cpsG"
6907. .8064
/ gene="cpsH"
/ codon_start=1
/ transl_table=11
/ product="capsular polysaccharide repeating-unit
polymerase CpsH"
/ protein_id="AAK11665.1"
/ db_xref="GI:13022166"
/ translation="MRKYLDLDYSLFYALWVLLVPNQMYQELIITIIVLLLLWKESEF
RISNSNSILFLMLFIYLFALLIRGTQEDITFQFIAELKLISITGYALFFYNYRK
ADFNSSVVRNVVKNVYFLVTLVLYLFPNEFTFLGDLFSIEWIPSMKVLRTAYF
EYATLIGOFILFTYPIFLKQOYGENIFITLFLVCAVLTGARFLICMITLLGVLL
LEIINKENLKITKKAPELLIIGIILLVCFYSKYVESINIYIHRFQSSSTRLLVYV
ESIRAILDGNFLIGOGIRVPSSVGIFLGSHSYIISIFYRTSTGLFLFLFLLFLYRE
AIKQNRITIKLFFGLLLYMFEEFDPNHSVLLFTTLGIVGRGNDKKTS"
8034. .8040
/ gene="cpsVII"
8034. .9013
/ gene="cpsVII"
8045. .9013
/ gene="cpsVII"
/ codon_start=1
/ transl_table=11
/ product="beta-1,3-glucosyltransferase CpsVII"
/ protein_id="AAK11666.1"
/ db_xref="GI:13022167"
/ translation="MIKKLVSVIVPVYNSELVIENCVESLLOQTYPEIEILLIDGSGT
DKSSHICNNFLKRDNRVKVYHKYNGGASSARNVGLMAEGEFTITVDSDDVALANNIE
IMLNLLTNADISEIDFEVSDDFYKRRKGYRVFVFNKSLKFEFFSGKNVENVVWG
KLYKSIIGDLFPNEKYIGDLEFNFOILANKHRIVVDTRBSLYTRYEEKSINQOQ
FNKNTLDTIDIFNEIHQDSPTLEFNVEAKFVREKIKCLRKNMPELGEIADENLRQRY
KFWQDIKSYISICKAIRLSKSHICTLYILMKYPPYVIKMYNKFQKQ"
9031. .9037
/ gene="cpsVIJ"
9031. .9752
/ gene="cpsVIJ"
9060. .9752
/ gene="cpsVIJ"
```

```
/codon_start=1
/transl_table=11
/product="beta-1,3-galactosyltransferase CpsVIJ"
/protein_id="AAK11667.1"
/db_xref="GI:13022168"
/translation="MSIYNPLKYIKSSIDSLISQILKNPEFIIVIDNPREDVDYIL
QTYLSDARVKVLSNERNGLAYSLNKAIRKYSVGNVYFRMDADDISYPERFQINFM
EKYRLDSATLIEVMDEDEETLYKORINKNFESDIYRLMSNRCVLSHPWCVKRAV
FLRLDGYRNVPAEDYDFLLRLALSGLIKGLIGLIDLLRYSNANGISQLININSIVTQ
LYPRIGSKWDKC"
RBS 9965..9971
gene /gene="cpsk"
9965..10913
```

```
Query Match 38.7%; Score 363.4; DB 1; Length 16448;
Best Local Similarity 62.4%; Pred. No. 8.5e-47;
Matches 585; Conservative 0; Mismatches 351; Indels 1; Gaps 1;
```

Qy	2	tgaattatagatcattatcggtgataatagagcctttaaaattgtgagagattcag	61
Db	9043	TGAAGTATAGTGCATCATGTCAATTTAATGAACCACTTGAAGTACATTAAATCTTCAA	9102
Qy	62	tagaattatataaaacaaacgcttactgatttggatttcataattgtcattgataac	121
Db	9103	TAGATTCAATTTATTCACAACAATPAAAAAATCTTGAATTTATTATTGTGATTGATAAC	9162
Qy	122	caagtagagtgatttaaagcaactcttaacagaattcgcgttggtagaataagaataa	181
Db	9163	CTGACAGAGAAGATGTATATGATATTTTGCNAACATATCTAGCAAGCGATGCAAGAGTAA	9222
Qy	182	aaatcttgcttaagaagaataattggtttgtagcatcaagtttgaacaaacgctgaaaa	241
Db	9223	AAGTATTATCAATGAGAGAAATATGGGATTAGCTTAGTTTAAACAAGCTATAAAT	9282
Qy	242	tttctaaggagaaatatattttagaatgagtgatgctgatgatttccatataccaagt	301
Db	9283	ATTTCAGTTGGAATATATTTTGAATGGATGCGGACGATATTTTCATATCCAGAAAGAT	9342
Qy	302	ttgataagcaaatcgttttatgaggaataattcattggatttctcagcaactctaata	361
Db	9343	TTGAAAGCAAAATTAACTTTATGGAAGATATCGTTTGTAGACTTCTCAGCACTTTAATAG	9402
Qy	362	aattgatagaccnaaaaggaatttagtatataaaacacgagaaagtaaaaaataact	421
Db	9403	AAGTTATGGATGAAGATGAAGAAACATTATATAAACAGAAATAATAAATAAATAATT	9462
Qy	422	taactaatgatatacgaagaatgttattgaatagatctatacttggccaccaccaacg	481
Db	9463	TTTCAGAGATATTTATAGAAATGCTCTCTAATAGATGTGTCCTTTACATCCCACTGGT	9522
Qy	482	gcgcataaagaagttttcgcataagttaatgggatagagatttagtacctgttgaag	541
Db	9523	GTGTTAAAGACGACGATTTTATAGGCTAGATGGATATAGAAACATTTGTTCTCGACAGG	9582
Qy	542	attatgattttgcaataagaggagctgctgctgatttccaaatcggttactcaataa	601
Db	9583	ATTATGATTTTTTACTAAGGTAGCTGTGTGTCAGGAATTAATGTGGTTTAAATCGTGATA	9642
Qy	602	tactttcacagtatagattaaacagagaatggaatatcacaaaccaataagttaa	661
Db	9643	TATTGCTTAGATAGGAGTAATGCTTAATGGTATTTCAC-AACTAATCAATATAAACAGT	9701
Qy	662	atattactcagctattttacaagatttttataaagaataatccttatttgatatac	721
Db	9702	ATATACGCAGCTATATCTCAGAGTTGGTTCAAAAATGGCAAAATGTTAAATGTAGACT	9761
Qy	722	aaattactaattacttcaagagtatgtgataaagaacgcgtactcagcagaagctct	781
Db	9762	ATATTCTTATTTATCTTGAAAGCTATATAAATAGCAATAAATTTGTCAAAAACACTTAA	9821
Qy	782	ctaaatatatttgagctaaataatcaccctcagtagtattactatttagaaaaactata	841
Db	9822	AAAAATACCTAGACTTTTAAATTTTAGTAGTAGTCTCTATGCTTAAAAAAAGTCTTTT	9881

Qy	842	tatatattatactttaagtcctcccttggttaggaggttattataataataatgattatt	901
Db	9882	GATTTTATTTTCAATATATCTTTTCACGTCGATTAATTAATCAAGATTTTATATTT	9941
Qy	902	tagtactagaaaattgttggagagagaaacaaagtga	938
Db	9942	ACAAAATAAAATATTGAACGGTAAGAGGACACAGA	9978

RESULT 3

LOCUS	U67549	14561 bp	DNA	linear	BCT 28-JAN-1998
DEFINITION	Methanococcus jannaschii section 91 of 150 of the complete genome.				
ACCESSION	U67549	L77117			
VERSION	U67549.1	GI:2826363			
KEYWORDS					
SOURCE	Methanococcus jannaschlii.				
ORGANISM	Methanococcus jannaschlii				
REFERENCE	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanococcaceae; Methanococcus.				
AUTHORS	1 (bases 1 to 14561) Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F., Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B., Borodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschlii Science 273 (5278), 1058-1073 (1996)				
JOURNAL	96337999				
MEDLINE	2 (bases 1 to 14561)				
REFERENCE	Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F., Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B., Borodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C. Direct Submission Submitted (27-AUG-1996) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On Jan 30, 1998 this sequence version replaced gi:1591709. Location/Qualifiers 1..14561 /organism="Methanococcus jannaschlii" /db_xref="taxon:2190" 184..1056 /gene="MJ1057" 184..1056 /gene="MJ1057" /note="similar to percent identity: 31.93; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glycosyl transferase, putative" /protein_id="AAB99061.1" /db_xref="GI:2826364" /translation="MVGEMDKPLVSVVMTATYNEPEKYLKESIESIXNQTKDFYIIVLDNPNKAEIEIKYQOKRKRIIFIKERNLGRGASRNKAVNIAKRYITAILDADDIALPKRLEKQFKYMNNRDIIDLLFSWYFIDENGNIKFPEKFKYKFFKEHLTVHPSMMVKSKILKLYKDEKILRSQDYDFWIRCIANDYKFDIEEFLKRYIPNRNLYLSRIKKOKLYSYVTLTWKHNKKHFCNNVYFWKVFVSLVVLIVLPTTFLIKLILDIKDKKTEISTKGH" 1067..2896 /gene="MJ1058" 1067..2896				

source

gene

CDS

gene

CDS

Db	221	TAATGGCAACATACACGACACCAAGAAAAATATATTTAAAGGAATCTATTGAGTCAATTTTAA	280
Qy	77	atcaaacgccttagctattgagttgcataatgtcatgtataatccaagtagagtgatt	136
Db	281	ATCAACATVTAAAGATTTTGAWTTTATAATCGTTTATAGATATCCAAATAATAAAAAAG	340
Qy	137	taaagcaattcttaacagaaattatcattagttgttagataatagaataaaaactcttgcattag	196
Db	341	CAGAGGAAATTAATAAGAAATATCAACAGAAAGATTAAGAGATTTATTTTATAAAAAATG	400
Qy	197	aagaaatatgtttagcatcaatgaattgaacaaagcgtgaaatatttctaaggagaat	256
Db	401	AAAGAAATTTAGTAGAGGAGCCAGTAGAANAATAAGCTGTTTANTATACAAAGGGGAAGT	460
Qy	257	atattttagaattgtagtgcgatgatattctatctcaccagtagatttgataagcaaatc	316
Db	461	ATATTGCCATATTAGACGCTGATGATATTGCATTACCTTAAGAGATTAGAAAAACAATTTA	520
Qy	317	gttttatgaggaaaattca----ttggattcttcgcgaactctaataagaattgatagacc	373
Db	521	AATATATGGAATAATAGAGACATTGATTATTGTTTACGTAGCTGGCTTTATTTATTATGATG	580
Qy	374	aaaaagaaatttagtatataaacaacagagaagaatataaaatatataacttaactaatgata	433
Db	581	AAATGGAATATF-----TTTAAAGAGTTTAAAGCCAGAAAAATATAAATTTAAAGAAA	634
Qy	434	tacggaagatgttattgaatagatctatctacttgcaccaccacgcgtgctgaataaaga	493
Db	635	TTAAAAATATTTTTCAAAGAACATTTAACTGTTTCATCCATCTATGATGTTAAAGAGTA	694
Qy	494	aagtttcgataaagtttaattgggatagatagattgacctgttgaaattatgatattgt	553
Db	695	AAATCTTAAAGAAATTTAAAAATACGATGAGAAATTAATTAGGCTCAAGACTACGATTTT	754
Qy	554	caataagaggagctcgtgctgatttcaaaatcggcttacttcaataaagaacttttaccagt	613
Db	755	GGATTAGATGTATAGCCATGATTACAAATTTGACATAATTGACAGAAATTTTATTAAAT	814
Qy	614	atagattaaacgagaatgaatatcacaaacaaataagtttaagcaatataattacttact	671
Db	815	ATAGATTCCAATAGACACAAATTTTAAAGCAGAAATCAAAAAACAAAAATTTATATTC	872
RESULT	4		
AF285774			
LOCUS		34508 bp	DNA linear BCT 15-JAN-2002
DEFINITION		Bacteroides fragilis NCTC 9343 PS B capsular polysaccharide	
ACCESSION		Biosynthesis locus, complete sequence.	
VERSION		AF285774	
KEYWORDS		AF285774.2 GI:18152896	
SOURCE			
ORGANISM		Bacteroides fragilis.	
		Bacteroides fragilis	
		Bacteria; CFb group; Bacteroidetes; Bacteroidales; Bacteroidaceae;	
REFERENCE			
AUTHORS		1 (bases 8470 to 34508)	
		Coyne,M.J., Kalka-Moll,W., Tzianabos,A.O., Kasper,D.L. and	
		Comstock,L.E	
TITLE		Bacteroides fragilis NCTC9343 Produces at Least Three Distinct	
		Capsular Polysaccharides: Cloning, Characterization, and	
		Reassignment of Polysaccharide B and C Biosynthesis Loci	
		Infect. Immun. 68 (11), 6176-6181 (2000)	
JOURNAL			
PUBMED		11035722	
REFERENCE			
AUTHORS		2 (bases 1 to 8469)	
		Coyne,M.J. and Comstock,L.E.	
TITLE		A Conserved Region of the Bacteroides fragilis Chromosome Upstream	
		of the Polysaccharide B Locus Contains Genes Implicated in the	
		Synthesis of GDP-L-fucose	
		Unpublished	
JOURNAL			
REFERENCE			
AUTHORS		3 (bases 8470 to 34508)	
		Coyne,M.J. and Comstock,L.E.	
TITLE		Direct Submission	
		Submitted (07-JUL-2000) Channing Laboratory, Brigham & Women's	
JOURNAL			

```

/product="Gmd"
/db_xref="GI:18152900"
/translation="MKKALISGCTGQDGSFLAEFLQKGYEHVGHILRRSSFNTRGRIE
HLYDEWVRDMKOKHTVNLHVGDWTDSSSLIRIIQQVQDEIYNLAOSHVKVSEFDP
EYTAADALGILRLMEAVRIILGLEKQTRIYQASSELGKQVQSEPTTFYPRSPY
GVAQGYWITTKNRYSESGYMAVNGILFNHESERGETFTVTKKTIILAAARIAQGEQDK
LYLGNLDAKRWGAKDYVECMVLILOHDFEDFVIATGEMHTVREFTCLAFAEIGIN
LRWEGEVNKGIDTATGKVLVEVDPKYFRPAEVEBQLLNGNPTKARTVLGWNPKCTPFP
ELVKIMVSHDMAKVRMIATKHD"
5702..6973
/note="Orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL61892.1"
/db_xref="GI:18152901"
/translation="MNDIFDELSSNSWNGKLPKVEFARKSYTNRFQYVGNRLVKVL
TGQGVGKGLRIIYQLIADGVSSNLIYSKFTDSGLSGTDRIDDDLLFIYREK
IHPKVGVYFIEEQKVEGWEHFHVSQDFTDCELFISGSDMLSGEAEKLLARH
YVSFEIFFSESEHRLERDEATGKNYAEYMGKALPILFNLPAEAGWKATISALRDT
ILFRDIIRRYIKDARLLEDVYVLTHLAELITIGQLVSHFAQNRKTSYDTVANYI
CYLEPISLHVERVOIRSKETLLGSKYIINDWSFMRHLYPFTAGSRKSCFENQVYL
ELRAGYTVFVGLRGCKLVDVGRKKDRIVYLOCAPLLNDDFQVEQMYNLEMLIQDN
YEKVVMSDDTFLPSKEGIRHIQVWLPEIL"
complement(7267..8070)
/note="Orf4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL61893.1"
/db_xref="GI:18152902"
/translation="MATTYDGINFPPIAAVNFMEVNAMEVIEAKYMGKGLAIVLKLLCK
IVKEYYIPWCEOCLIFANKTGKEADAEVQGIIEIMIEKGMIDRGSYAKHKVLTSE
AIQKWIETARKERNWTFAMPYLLIKPKETCNKCTQNVQEQDAGSDAKNACNTEQ
SKYKOSREPPSAPRGKEEVNATPVSNPQTAFNTWNTVYVGLMGNLERFGITDEKE
IEAILRLSDYGRPPKVPWKLICSTNWSNIGAKGYMIAALNRAKRSQT"
complement(8120..8467)
/note="Orf5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL61894.1"
/db_xref="GI:18152903"
/translation="MKKNRKAVSKNEMMKWKWKQIMLEKDYTEQCSEWMAERLEALI
EYMQYGHAAVAYMKQDGTFLVKVGTLVGYEKDFGQYDPMPEIKNTVYVRDVEQQRWMT
FKIENFEWRAIV"
9468..9992
/gene="ubpY"
9468..9992
/gene="ubpY"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulatory protein"
/protein_id="AAG26461.1"
/db_xref="GI:11023504"
/translation="MSEQQKYWFAARTDRDKQFEFAIRDSLEKLKLTDLNYYLPTQFVI
ROLKYRRKRVVPIKILNFIQATKQDADCSNKYNIOLFYMKDLITRAMLIVDPKQM
QDFIVMDLPNGVSFDNDHLSVGSRVQVVKGDFGCVGESELANEANKYVIVIRIAGVL
SASVKVPKSYLRVI"
9996..10481
/gene="ubpZ"
9996..10481
/gene="ubpZ"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG26462.1"
/db_xref="GI:11023505"
/translation="MNTLTSTQEQLSLAHELLYLGVDGAPLYTDHFRQLNKEVLEQS
DALYQRCATPEEENACILALMGYNATYINOGDKEEKQVVLNRCWDLVDQLPATLL
KCOLLYTCYGEVFEELAKEAHTIIESWNSRELLKAKEIAESLNLEANPYPSSELH
E"

```

```

gene
10478..11761
/gene="wcfT"
10478..11761
/gene="wcfT"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG26463.1"
/db_xref="GI:11023506"
/translation="MIRFSLRSPLYLALYASVLIFLVYTCAYARKPFTAAIYEGE
ILMGFDVKILYLSLIGYALSKFIVGILPSMKAGHRIYVYIGLITSEVALLGFLAL
LPVPLKVCISIFLSPGTLGIMWVIFSYTEGRRISEILNGLSVLIVSGLVKTLGQFQ
VMDNLHVSEYWMPTTGLLVYPVMDLISWLLNQIPAPNEQDIIQRTKRAPMDKDRRK
FLRQFFWICMLVVFYGTALTVELRDSFAADINWELHIEGAMIFQTQTEPVTAAFLV
LMFLIVFVRNRLALNIYCIATVGGSLMVFSTLLYVYGFSLSPIMWMLISGLGLYMGY
IPFTYLIERLIASLRVSTAVFIYVLADSFVGLTGTGVFMVKNFVSADISWTVMLMRT
ILFSGFVSVLSIFAIYCYCFKQQLNSLILPSEIHD"
11754..12473
/gene="wcfU"
11754..12473
/gene="wcfU"
/codon_start=1
/transl_table=11
/product="putative glucose-1-P-cytidylyltransferase"
/protein_id="AAG26464.1"
/db_xref="GI:11023507"

```

```

Query Match      12.9%; Score 120.8; DB 1; Length 34508;
Best Local Similarity 60.2%; Pred. No. 5.8e-10;
Matches 200; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 4 aattatgattatcattatcggtatataatagagcccttaataattatgagagattcagta 63
   || || || || || || || || || || || || || || || || || || || || ||
Db 22262 AAGATTAGCGTATTGATGCTATATATAGAGAAATCTCTTTATGTAAGACAAGCTATT 22321

QY 64 gaattatattaaatcaaacgcttactgatttggtgattcatttgctattgataatcca 123
   || || || || || || || || || || || || || || || || || || || || ||
Db 22322 GATAGTATTATTAATCAAAAGTTTGGAGATTTTGAATTTTAAATATATAATGATAATCCA 22381

QY 124 agtagaggtgatttaagaacattcttaacagaataattcattgattgataataaataaa 183
   || || || || || || || || || || || || || || || || || || || || ||
Db 22382 GAAGATTTTATTTAGATAAGTTAGTTTGAATAATAGCAAGAAAGATAGTCGTATACAA 22441

QY 184 atcttgcttaagaagaaataattggtttagcatcaagtttgaacaaacgggtgaaaaatt 243
   | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 22442 TATCAACGAAATGAAAAAATATAGGTTTAGCAGGCACATTGAAATCGTCAGATTTTGCTA 22501

QY 244 tctaagggagaataatatttttagaattggatgctgatgatatttcataatccaagtagattt 303
   ||||| || || || || || || || || || || || || || || || || || || || ||
Db 22502 GCTAAGGAGTATTTAGTCGTATGATGATGATGATATTTTCAATTCACCGATAGATTT 22561

QY 304 gataagcaaatctcgtttttatggaggagaaatttc 335
   || || || || || || || || || || || || || || || || || || || || ||
Db 22562 ATTATGCAGCTTGAACTACTAGATAGAAATAC 22593

```

```

RESULT 5
AE007717/c
LOCUS Clostridium acetobutylicum ATCC824 section 205 of 356 of the
DEFINITION complete genome.
ACCESSION AE007717 AE001437
VERSION AE007717.1 GI:15025159
KEYWORDS
SOURCE Clostridium acetobutylicum.
ORGANISM Clostridium acetobutylicum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
REFERENCE 1 (bases 1 to 14157)
AUTHORS Nolling,J., Breton,G., Omeichenko,M.V., Markarova,K.S., Zeng,Q.,
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucalle,P.,
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.

```

TITLE	Genome Sequence and Comparative Analysis of the Solvent-Producing
JOURNAL	Bacterium Clostridium acetobutylicum
MEDLINE	J. Bacteriol. 183 (16), 4823-4838 (2001)
PUBMED	21359325
REFERENCE	11466286
AUTHORS	2 (bases 1 to 14157)
TITLE	Childress,D., Zeng,Q. and Smith,D.R.
JOURNAL	Direct Submlsion
FEATURES	Submitted (24-JUL-2001) CTC Sequencing Center Production, 100
source	Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
	Beaver Street, Waltham, MA 02453-8443, USA
	Location/Qualifiers
	1. .14157
	/organism="Clostridium acetobutylicum"
	/strain="ATCC 824"
	/db_xref="ATCC:824"
	/db_xref="taxon:1488"
gene	complement(138. .1136)
CDS	/gene="CAC2166"
	complement(138. .1136)
	/gene="CAC2166"
	/codon_start=1
	/transl_table=11
	/product="Nucleoside-diphosphate-sugar epimerase"
	/protein_id="AAK80124.1"
	/db_xref="GI:15025160"
	/translation="MNILVTGGAGFIGRWVVKLLDDGNTVTALDNLNNGSIENIEEF
	KGNFPIEGDIKDDLDRIEDKYDIIVHLAASIVQESIDDPEKTFENDVTGTFN
	ILERAKQPGQHGKNGSINWIDENDEYYPCKVFWPMSCTWYVDISGETIDQHSVK
	PVSPYGSKIAAENMVLVSFYNAIKLPAVIRPENTYGPFOKSGEGGVSIPLKNFID
	NKPINTYSGSKOTRDLLFVKDCADFYVNAKYSEVDEGIINAGTGNDITINELANLIA
	GKVSYNHVEHIHPQSEIKMLKCNYSKAKNLGMWEPHYTLEEGIAETKSWLASTIK"
gene	complement(1268. .2119)
CDS	/gene="CAC2167"
	complement(1268. .2119)
	/gene="CAC2167"
	/codon_start=1
	/transl_table=11
	/product="Flagellin family protein"
	/protein_id="AAK80125.1"
	/db_xref="GI:15025161"
	/translation="MRINHMAALSIFYEQKSVGGQOATSLERISSGLKTNSAKDGA
	ALBKSQMRQIIRGMAARNVQDGIISMFQAEAGMSNVSDMLQMRDLVLESGDSK
	TSDKNIQIIEIKINGINDITNTNFNGKPLLASNEKSEITIGNNGVTFQSFYD
	LSSESGOKDANGNVTSIRNVNLQADSLANLATDSALSTVLDLRAOXYENRPE
	DTYNLNSPESGVQAESDIRDSVADENITYTKTGVLMNAGDAMLAAQNLPKDILN
	ILSSIRK"
gene	complement(2227. .3963)
CDS	/gene="CAC2168"
	complement(2227. .3963)
	/gene="CAC2168"
	/codon_start=1
	/transl_table=11
	/product="Uncharacterized conserved protein"
	/protein_id="AAK80126.1"
	/db_xref="GI:15025162"
	/translation="MNIITFNTVWGCRDISEINAVIEESLDNKLKLDNKGNNREN
	VQEDINSFKKLNIENETESDFVIFGLCGEHLKLLNMNLKMKRLILVEPREDVISRF
	LKLYANKIIDDRIVLYLYNEEDNLGILSLFDENISNTKCGVFANYADTYEELS
	NLYKSIDIONSLLVNMGTHVHSRHFSSYMMNLKFIQKSTIINHFNKIYKGPVAVI
	VSAGPSLNIDLDKYQSKFIVITGGRTKLKLLDGIKRPDFVCVIDDPEAFEIMKD
	ALDCTVP IYSEFTNKKVINKYKDKIFFDTGVDNASDKDFFEKIDNLYEGGSVAHV
	CASLAEY IGCNRIIFYQDFAFYNDVTCDLNVGNSNIKNFYVDVYGGVKVKTDRTL
	YFYKMAEEFIQVKKHVDFTNCTEGGANIEGAPLIKPKDALEKGESEIKNTEAVSNLI
	KSNVLMDDNMRIKLDGIRQNIKVISSMCTSAVEQYNSYLAFGKSESOKFLAHNF
	NLNFECIFSNASSFKFINVLSPAILNIGLNPREKMGKMBEKKIKIKERNLLIYN
	NITEAAKEFNKY"
gene	complement(4024. .5106)
CDS	/gene="CAC2169"
	complement(4024. .5106)
	/gene="CAC2169"
	/codon_start=1
	/transl_table=11
	/product="UDP-galactopyranose mutase"
	/protein_id="AAK80127.1"
	/db_xref="GI:15025163"
	/translation="MEDYLIVGCGFCASARLLAEXGKKLIVDKRHHGGNAYDIY
	DLNGILVHKYGHIFHTNSRKRWVEFLSRFTKWLVIQHKVLSYDGRVLVPPIMPVDTVN
	QLYGTNYSNDTI SEFFDKREKDIENARDMVYTKVGSOLYKMFYTKKONGIMP
	EELKEVETARIPITNRDRDYFTNYOGLDLDGTYKMFENMLNHNTHMLNTYKFI
	KSEIQKSLIYTGCADEFDYKFGRLPYRSINFETLNODYVOKGVTVNYPNDYDFT
	RITFEKHLTGKCYNTTIVREYSCSDGEYYPVPQSKNQKLYLEYKQVENLENAYFL
	GRUGQRYMNMVVDRAIKLVDKLK"
	complement(5153. .5638)
	/gene="CAC2170"
	complement(5153. .5638)
	/gene="CAC2170"
	/codon_start=1
	/transl_table=11
	/product="Predicted membrane protein"
	/protein_id="AAK80128.1"
	/db_xref="GI:15025164"
	/translation="MKRLIVRSASFQOOLDNLKAIKKYKYDEIDILTHEHGVLAE
	KYNNIHVYVPYKSEPKGKVKKEINIKFDVVIPTVNTISGAGFNVPKYALSINS
	DKIIMCNVISEFKDRDRLSIYIMELSSFVKILSGFTCLFGIVSVFLIFLKLRAAK
	K"
	complement(5661. .6560)
	/gene="CAC2171"
	complement(5661. .6560)
	/gene="CAC2171"
	/codon_start=1
	/transl_table=11
	/product="Predicted glycosyltransferase"
	/protein_id="AAK80129.1"
	/db_xref="GI:15025165"
	/translation="METPLYSIVSICWNKRDEICESLNNIRKIDYESLEIIVDNGST
	DCTVEKIEEDPKVKLLKMPKNAGIEAYNVGFKSAAGEYIVILDDDFPEKNSIKRW
	KKQKDKLGMVAFDRNRYNDYDKADITEETNSDAAESDEYIMAFGAGAVRTE
	VLKKAQYEEFFLYWNEQDTAFRLDMGYKIKFFADVIAHYKSPKRNASWRAPFY
	TRNAFFLVKNYSTKAALKRTIELIYDCFYYSMEQKNTIYLKAMWNAFMNFNKKIGKR
	KVVDKYIENLRIPLVNSFTFYK"
	complement(6689. .8038)
	/gene="CAC2172"
	complement(6689. .8038)
	/gene="CAC2172"
	/codon_start=1
	/transl_table=11
	/product="Predicted glycosyltransferase"
	/protein_id="AAK80130.1"
	/db_xref="GI:15025166"
	/translation="MKELSIVCNKNKEYVLNLCIDSVLSSYRDLDIYVVVDNASEDG
	SVEAIKDKFKDVKMLISNTENIGSGGFTGIREVLKDKYKVMLLDNDVILDKKAIN
	NSYRYLEKNDOTAVVGSKIYSMDNRDKIQEMGAKIDFENFYINPFYGYIDFIELPEV
	VEGDYVPACSMVRIEALRRVIGIMEDDNFIWDDIENGWYRFKAGYKVTYGGKVVWH
	KMGVAOKTNTFGTYFWNRNIRPEAKYIENRDKDFCLRFDEIQAITYSCNYIGKYS
	SARTIIAAVEDALNDVRGKASQDKILIEIKVKDRFGLIENKKNKIALTGDSNIEVLN
	IVNRIKATNSETELNLSGNVKNIKQOFEYKVLKSESLKTYETVIKICNHFDMR
	NELSNLIYVDKFPNVIVSGDDRKYIENYDNTYSLMKNTIYNLLSKRMVLRNLKLNK"
	complement(8063. .9064)
	/gene="CAC2173"
	complement(8063. .9064)
	/gene="CAC2173"
	/codon_start=1
	/transl_table=11
	/product="Glycosyltransferase"
	/protein_id="AAK80131.1"
	/db_xref="GI:15025167"
	/translation="MAKVSVMPVYNSEKYLKESIESILNQSYSDLEFIINDGSTDN
	SFKTIKLYAKLDRKRNVISRENKGIYSLNAEIRLAKGEYIARDADDISAPKRIEQ
	ISFLKSHRDDILGTQVKNISNDIKENKNLNIEFDIYDNRNREKILNLYWYCLAH
	PSVMFRKIDILRELQYNDKFSDELDLWLRATESGFKYIKLEEILYFMHESKTRVD
	NONYEGLKDGIKIKLIDVKEKREFKDFKIVWVGASNGKIKTKVELDDEFFEKSOIAFY
	DKFTGFEKIRILHPKIDINHIFDYVFIATEPKCKEAMSSLSKGLKCKIKDFLSTV"
	complement(9115. .10125)
	/gene="CAC2174"
	complement(9115. .10125)

```
/gene="CAC2174"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK80132.1"
/db_xref="GI:15025168"
translation="MKVYIPKVSIVMPVNSERYLAETIESILDQTYNDFEPIIYDDGS
TDESNIILSSVANKDNRIIVISREHGLVDSLNEGIDRAKGSIVARMDDADISINNRI
EKQFELKNDVDILGTRIEAGDIDKQRTIYNSAFSIFKSDONIQEVLFTSCAIP
HPSVMFKKSDSVKLGRYKEDYADYDLMLRINRGYKIVRMDCLIKYRVHNSKT
AVENPMKWVEYTKAKIDYINDTNKDKVDYLWGASTGKLVKKVYESTDKFNLI
GFIDYKEGEVDECRYKPEDLKMKVDYVFATLPGKEADEFLKNCGLKFKFEYIS
Lv"
gene
complement(10149..11150)
/gene="CAC2175"
complement(10149..11150)
/gene="CAC2175"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK80133.1"
/db_xref="GI:15025169"
translation="MISVIMPVNCEKYLESIIESILKQYRDEPIIYDGSNDKSI
DILNKANDNRIVVSRDNNMGWYISNEGIDRAKGSIVARMDDADISINNRI
EYLNKNDVDILACKVEAFDVSREQKLEREHWNVDLNNSESIESTLENCYTAHPS
VMKMSVLKALGGYNLNRKTEDYNLWRAIAKGYKIAMLEEKLMKIRLHNSDKTHRD
AEGSSIRDIITQSRLEYKELKLDYSYIWGASNGGKIAYEKIKEYFPNAKNGYI
DKFTGKPEGVDFIPYEDLSKMKVDYVFATLPGCKEAKGILTKNFPVINKDFLVC"
gene
complement(11189..14002)
/gene="CAC2176"
complement(11189..14002)
/gene="CAC2176"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK80134.1"
/db_xref="GI:15025170"
translation="MIKQDFVSDFSIERSNRLIESPNVSIILCLINIKDYNKLDKSI
SLILNQYRNFELIVDDQDBELTYTKIDYINKDNRIVYVHNRRSGLAALRINQA
Query Match
Best Local Similarity 50.1%; Pred. No. 1.9e-09;
Matches 467; Conservative 0; Mismatches 440; Indels 25; Gaps 6;
QY 13 atcattatgcgtatataatgagcccttaaatatgtgagagattcagtagaatacata 72
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9052 ACGGTGTGTAATGCCAGTTTATAATAGTGAAATAATCTCGAAAGAGTCCATAGAAAGTAT 8993
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 73 ttaatacaaacgcttactgatttgcgttcataatgttcattgataatcccaagtagaggt 132
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8992 TTAATCAATCATACAGTGTATTGGAATTTATAATAATAATCAATCATGCCTC---TACAGAT 8936
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 133 gatttaaacgaattcttaacagaattcagttgtgataataagaataaaattcttgcct 192
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8935 AACTCTTTTAAATAATTAAAGGAATATGCCAATTAGATAAAGAAATAAA---TGTGATT 8879
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 193 aatgaagaaattgtgttagcatacaagtttgaacaaagcgggtgaataatttctaaggga 252
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8878 TCAAGAGAAATAAAGGAATTGTGTACTCGTTAAATGAAGCTATACGATTGGCAMAAGCA 8819
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 253 gaatatatttttagaatgctgatgatatttccatcccaagtagatttgataagcaa 312
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8818 GAATATATAGCAAGGATGCTGATGATATTTCGCGCACCCCAAAAGAAATAGAAAAACAA 8759
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 313 attcgttttcatgga---ggaataattcattgatttctcagcaactctaataagaattgata 369
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8758 ATTTCTTTCTTAAAAAGCCACAGAGATATAGATATACTGGGAACACAGCTAAAGTGGTG 8699
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 370 gaccacaaaggaatttagtatataaacacagagaagaataaaaa-----ta 417
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8698 GCGAATATTCTTAATGATATAAAGAAAAAATAAGCTAAATATAGAAATTGAT 8639
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 418 tacttaactaatgatatacaggaagatgttattgaatagatctatacttgcacccaccaag 477
```

```
Db 8638 ATTTATGATGATAAATAGAGAAAAAATATTAAATTTATTTGTTAGTATTTAGCACATCCTTCA 8579
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 478 tgggtgcgtaaaaaaagaagtcttcgataagtttaattgggatatagagatttagtacctgtt 537
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8578 GTAACTGTTTAGGNAACAGATATATTGAGAGAGCTAAAAGGATACAAATGATTTTAAATC---A 8522
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 538 gaagattatgattttcgaataagagagagctctggctgatttcaaaaatcgcttactcaat 597
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8521 GAGGATTTAGATTATATGCTTAAGAGAGCTATAGAAAGCGGATTTAAAATATATATAAGCTAAAG 8462
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 598 aaagtacttttcagtatagattaaacgagaatcggaatatcacaaacccaataagtttaag 657
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8461 GAAGAACTTTATATATTTTAGGATGCGAGAGAGCTCAAGAGCTAGAGTAGATAATAATAAAT 8402
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 658 caatatatttactcagctattttacaagatttttataaagaaaaatcttatattgataac 717
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8401 TTGGAAGGTTTAAAGATGAATAAAAAATAAAATTAATAGATGTTTTTAAAGAGAGAGTTC 8342
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 718 acaaaaattactaattacttctcaagagatgtgataaagaacgctatactcagcgaag 777
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8341 AAGAAAGACTTTAAATACATAGTATGCGGTGCAAGTAATGGAGGCAAAATTAACAAAAGAG 8282
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 778 ctctctaaatattttgagctaaatctacccctagtagtatttacttacttagaaactatatatt 837
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8281 GTTTAGACGAATTTTTTTGAGAAATCTCAATGCTATTTGTCATTTCTAGATAAAATTTAAACA 8222
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 838 tgtttatatttacttacttaagctctcc-cttggttagaggttatttaataatgatattaa 896
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8221 GGTGAATTTGMAAAATTAAGATTTTCATCCAAAAGATATAAATCATATTAATTTGAT 8162
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 897 tattttagtagtaaaattgtttggaggagaga 928
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8161 TATGTTTTTATTGCAACTGAACCGGAAAAAGA 8130
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 6
AF279134 3584 bp DNA linear INV 05-SEP-2001
LOCUS Dictyostellium discoideum alpha1,2-fucosyltransferase (fuch) gene,
DEFINITION complete cds; and unknown genes.
ACCESSION AF279134
VERSION AF279134.1 GI:9022424
KEYWORDS Dictyostellium discoideum
SOURCE Dictyostellium discoideum
ORGANISM Dictyostellium discoideum
REFERENCE 1 (bases 1 to 3584)
AUTHORS van der Wel,H., Morris,H.R., Panico,M., Paxton,T., North,S.J.,
Dell,A., Thomson,J.M. and West,C.M.
TITLE A Non-Golgi alpha 1,2-Fucosyltransferase That Modifies Skp1 in the
Cytoplasm of Dictyostellium
J. Biol. Chem. 276 (36), 33952-33963 (2001)
JOURNAL 21423935
MEDLINE PUBMED
REFERENCE 2 (bases 1 to 3584)
AUTHORS van der Wel,H. and West,C.M.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2000) Anatomy & Cell Biology, University of
Florida College of Medicine, 1600 SW Archer Road, Gainesville, FL
32610-0235, USA
FEATURES
source
1..3584
Location/Qualifiers
/organism="Dictyostellium discoideum"
/db_xref="taxon:44689"
<1..>807
/product="unknown"
<1..807
/codon_start=1
/product="unknown"
/protein_id="AAF82377.1"
/db_xref="GI:9022425"
/translation="DLKQFDLSLNQYCDENILSTYLIGNACPRWTPPHQGHNRNNSQ
```

QQLYPSANPQOOLPPTILSQSPQKYTPPTNNHHHHHLLNPLSPRYSNTLSPTQOP
INTSVNNNNIIQORNSISSSNCINSISSIRRSFHDNIINGNGNNGVNVGSSSSLS
NTTSAFSPNVSYSGNSGSCNISSSMSMMMSGNOQPLIGSPSEMIHKLKE
IERLNKEATQKQDELQNVKIAEABERSRLCLKDKYKEYRNELIERELVK"
<1067..>3373

mRNA

/gene="fucB"

/product="alphanal, 2-fucosyltransferase"

<1067..>3373

gene

/gene="fucB"

1067..3373

CDS

/gene="fucB"

/note="glycosyltransferase"

/codon_start=1

/product="alphanal, 2-fucosyltransferase"

/protein_id="AAF82378.1"

/db_xref="GI:9022426"

/translation="MNDSPISVLPFLIKNDKSLNYQGINLLISIDSIIIEQTFK
ENELIVDDGSNEIEQLSKRYSDNRKFIINKENGIYKSLNDALINHCSPFSK
YIARMSDDISHPTQLQSKYLQSNETIDILGCPKMFNNKLLIEILNNNNNNIN
NAVKELDIIINNESEKFIQHPDKDILMWSMFNCCIVHPSVIFKRSIFTIEHCYEN
NOPTFEDLYFLKSLIMKGLNISQSTPLLYLRKHNNISFKNIEKOKSTANAS
CYVLTLPKREINDSEIIIONSLSMKEIIQFOLSPSSLSKINNIIEIPEPAFKYLE
LLEKSTCKQOPVYSNISKDAANEKKGELYSLSLSNYPNNQKSSLLMEKRLSRPTSQL
LSLLSLNVSSTTIINNINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLLNFIGIN
SNKINTPPKNNNKKFKENGIRIICFSKDRAPQLKEYLRTFFKYLKNDNDGNDKPEIIVD
VLTFTYSEKFKNSQVIEFQVNFPIKEENFTDOLINLVQKTNKLEYVMFSYDIIILY
YNEFLNKEYCLSLNSPLALGFQVFNKLNKNTYCHTCNQDITPLNSNTISRTENNFKY
LKNWRNDCKDKNVYDLCSTIYRCNDIDISTINGIVKYGIRNGINPNRFEENGN
RELIOKQIYQNKPYCCLSDHSPNSVNTINRVQDYVDNPIDYDTLSLDDLDLQLYSN
complement(<3471..>3584)
/product="unknown"

mRNA

/product="unknown"

/codon_start=1

/product="unknown"

/protein_id="AAF82379.1"

/db_xref="GI:9022427"

/translation="DPTKENNTVGHVYKGYRLUHLRLVRPVMGVNKRKP"

BASE COUNT 1487 a 442 c 365 g 1290 t

ORIGIN

Query Match 10.9%; Score 102.2; DB 3; Length 3584;
Best Local Similarity 50.3%; Pred. No. 6.7e-07;
Matches 285; Conservative 0; Mismatches 273; Indels 9; Gaps 1;

Qy 1 atgaattatagtcattatgcggtataataagagccttaataattatgcagagattca 60
Db 1109 ATAAGGACAAATGACGATAATCAATTAATTAACCAAGCAATTAATAATTAATCA 1168
Qy 61 gtagaattctattataacaaagccttactgattttgatttcataatgtcattgataat 120
Db 1169 ATAGATAGCATTTTGAACAACTTTTAAGAATGGGAATTAATTTAGTTGATGATGGA 1228
Qy 121 ccaagtagaggtgatttaaagcaattcttaacagaattattcagttgttagataatgaata 180
Db 1229 TCAATAATGAAATTTGGGCAATTAATCTTCAAAAAGATATAGTACAGATAATAGAATT 1288
Qy 181 aaaaattctgtaagagaaaattatggttttagcatcaatgattgaacaaagcggtgaaa 240
Db 1289 AAATTCATAATAAAGAGAAATTAAGGATTTCTTAAAAATTTAAATGATGCAATTTTA 1348
Qy 241 att-----tctaagggagataatttttagaagtgatgcgtgatgatttcattat 291
Db 1349 AATCATTTGTCACCAACTTCAAAATATATCTCGTATGGATTCAGATGATATTTCTCAT 1408
Qy 292 ccaagtagattgtaagcaaatcgttttatgaggagaaaattcattcgttatttcagca 351
Db 1409 CCAACAAGATTACAATCTCAACTTAAATATCTTCAATCAATGAACAAATGATATATTA 1468
Qy 352 actctaataagattgatagaccaaaaggaattatagatatataaacaacagagaagaat 411
Db 1469 GGTGTGCCAATTAATAATGTTTAAATAATAATAATTAATGAAATTTTAAATAATAATA 1528

TITLE Analysis of the genome of an alkaliphilic Bacillus strain from an industrial point of view
JOURNAL Extremophiles 4 (2), 99-108 (2000)
MEDLINE 20263314
REFERENCE 9 (sites)
AUTHORS Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T., Hirama,C., Fuji,F. and Takami.H.
TITLE Characterization and comparative study of the rrn operons of alkaliphilic Bacillus halodurans C-125
JOURNAL Extremophiles 4 (4), 209-214 (2000)
MEDLINE 20426005
REFERENCE 10 (sites)
AUTHORS Takami.H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N., Fuji,F., Hirama,C., Nakamura,Y., Ogasawara.N., Kuhara,S. and Horikoshi,K.
TITLE Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis
JOURNAL Nucleic Acids Res. 28 (21), 4317-4331 (2000)
MEDLINE 20512582
REFERENCE 11 (bases 1 to 303650)
AUTHORS Takami.H. and Takaki.Y.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group; 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan
(E-mail:takami@jamstec.go.jp,
URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,
Tel:81-468-67-3895, Fax:81-468-66-6364)
FEATURES
source Location/Qualifiers
1..303650
/organism="Bacillus halodurans"
/db_xref="taxon:86665"
/note="alkaliphile"
complement(103..1887)
/gene="BH3486"
complement(103..1887)
/gene="BH3486"
/codon_start=1
/transl_table=11
/product="butyryl-CoA dehydrogenase"
protein_id="BAB07205.1"
/db_xref="GI:10176110"
translation="MGNTAQTIVKGGSLDEIDADKVFTPEDTFEEHEMIGKTTEF
VNEVPVLDIEIHQFDISRRLTQAGELGLGADVPEYGGGLDKDISSIITEKF
SRAGFSLSYGAHVGLSGLPVVFENEDQKKYLPDLATGARIAAYALTEPSSGSDAL
GAKTAVLINEAGTHVLNGEKWITNSAFADVFIYAKIDGHEFSAFIVEKDFEGVST
GPEKKMGKSGSTFTLLEDALVPKENLGEVKGHVIAFNILNVGYKLGVCIGG
SKRGIELAAKNAERKQKQPIAKFTLQEKLANWAVTYAAESSIYKGTGLFEDRLG
GLSEEQKDGREVAKAIAEYALECSLNKVGSEALDFVADEAVQIHGGYGFMAEYVE
RMYRDSRINRIFEGTNEINRLVPTILRKAMGELPELEKATLQELMLMPOEVG
DEPLEQERYLLSMARKVFLMIAGTGAQTYGKLLQEQELLANVADIVSDIFSMESVIL
RTERAIRNGLEAKQKALQVFCQEAFNIEAHAKESLSVAMQSGDPLRTWTSILRK
LTRHTPINVIKKREIAGRIEAERYV"
complement(1909..3090)
/gene="BH3487"
complement(1909..3090)
/gene="BH3487"
/codon_start=1
/transl_table=11
/product="acetyl-CoA C-acetyltransferase"
protein_id="BAB07206.1"
/db_xref="GI:10176111"
translation="MREAVIVAGARTPVGKAKNGTLANVRPDDLGAITVKETLRAGD
FDPARIDDIIVGCAPEAESMSLIIPMGGHVIAFNPTLVENQPEYVMGMHTAEVASRFG
RIMLQGEALIVAGAESMSLIIPMGGHVIAFNPTLVENQPEYVMGMHTAEVASRFG
VSREDQAFAVESHKRAAKAIEEGKFADEIYVPDVTLRHVDNHLKVEKHVTFSDVEG
VRPEINMETLSTLRPAFHAKGTVTAGNASQMSDGAVALMDREVAEREGLTVPVKFR
SFAVAGAPEIMGVGPVEAIPKAVMAGLDLSDGLFELNEAFASQATQVIGTLRLDH
DKVNVNGAIALGHPGLGCTGKTLTSLIHMKRRGEQFGVVVTWCTIGCGMGAAGVFELI
G"
complement(3111..5492)
/gene="BH3488"
complement(3111..5492)

/gene="BH3488"
/codon_start=1
/transl_table=11
/product="3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase"
protein_id="BAB07207.1"
/db_xref="GI:10176112"
translation="MARQIRKVAVIGSGMGSGIAAHLANVGIPSLLLDIYVNPQLTEE
BORKGLTLEDAVRNRLATLALKLOKOKAPLRSKSHVHYIQGNLEDDQVEKLAEDV
WIEVVVENLDVKKOLFVKDYQKQGTLVSSNTSGISIAMAKGRSDFQAHFLGTH
FNPFRYLKLEIIPTKTKPEVTAFAKFAEDVLGKGVVEGKDTFNFIAKRIYIGIL
LVTQKMLEGYSVEYDVSVTGPIGRPKSATFRTLDVGLDTFLHVAKNHVDVVDGE
EDVFDPFVFMKMAENGWIGSKQGFFLKKQKGENGSEILELNPSTLEYEKRRKKKA
PSIQAKQSLNKEKLSLIIANDRAGOLLNLMKPLVLIYSAEKYETAGDIKAVDDA
MKWGFQWBOGPFEMWDAGVASSVORMEEGATVPWKEMLTSGRERFYODCTFYHH
GEAKVIPNGKIHLKALKPERVITKNSGASLIDLDGDOVALLFHSPPNSIGFDVVOV
INKAVDEYANYKGLVIGNQKFCVGNANLMLMEADNDFEIDMVMYRFOQAMAK
IRYSOKPVASPAFTMGGTEICLPTASIQASLETYMLGVEVGVIYGGGNGKELY
LRHLERLPEGANIDQSTARKKIPVVGTYATMLLGAKSMKFGGMISEHDLKIAEKL
YQAKQQVILHDAGYRPPARKKIPVVGTYATMLLGAKSMKFGGMISEHDLKIAEKL
AFVIAGGRVAKGTLVDEQYLLDLEREAFSLVGEPKSQORMQHMMLMKGLRN"
5746..6015
/gene="BH3489"
5746..6015
/gene="BH3489"
/note="BH3489"
unknown
/codon_start=1
/transl_table=11
protein_id="BAB07208.1"
/db_xref="GI:10176113"
translation="MQQSFMQSGQQQGVMPQPPQVITVKDQLYITDMLSWNLLACKK
AHFTAQOCQDPEVKQALQEQAGQMQRHYTMLLGLHQQNNMATVQ"
6030..6353
/gene="BH3490"
6030..6353
/gene="BH3490"
/note="BH3490"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
protein_id="BAB07209.1"
/db_xref="GI:10176114"
translation="MNQQKIQNPETPVQKNTQMSDRDFINDOLATEKYITSSYIAM
NEASHQOLYQDIFACQEAQDCORQOLYELMFKGWYSELAAPQKLEAQYQOFGYMN
QLPYH"
complement(6396..6644)
/gene="BH3491"
complement(6396..6644)
/gene="BH3491"
/note="BH3491"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
protein_id="BAB07210.1"
/db_xref="GI:10176115"
translation="MDEKLQQLDGLVEKYAELLVGDTSSEMKERIRAYIYSHISK
MPPLAKHWSQVPPAKEGMKALFLDIQNLKKQRENKQ"
6761..7150
/gene="BH3492"
6761..7150
/gene="BH3492"
/codon_start=1
/transl_table=11
product="transcriptional regulator (GntR family)"
protein_id="BAB07211.1"
/db_xref="GI:10176116"
translation="MNYTFDDSRPLFQQVADAIANDIHNVIKEGEQVPTNQFAKHH
QINPATVAKGHLVLDQGLFKRGVGMFVAKGARDQLLEQRRETFYSDFIYPLLINEA
KTLHLSSEIMLDLIKRGANDSDDHM"

Query Match 9.7%; Score 91.4; DB 1; Length 303650;
Best Local Similarity 50.8%; Pred. No. 1e-05;


```

NIRSMILFGNEYNPLIYKYITLQSELLNDLSTIEHDMKYIINDDHNLKSGQKVRICLA
RALYEHYIHMILKCTDYDEKKLIQNPNEIIDLKOLINLKNISSYNNKSKVLNYPINPEN
YLOKCLMDNNPFYLLDDIPTSLDPSISKIFSNLKCKEONISPKDNCSTFISMNKS
TLDNPLIEDILNDVOYEVNIELODKTLKYRGNISEYMEKANLNTIKESHGYSNLNT
IDYTRIKLEDELNVHNSKMYIKEYAVYKWNVESFEDISINKIYIKMKKKNY
KHEHNNKONNNNSNDHININNDONHRNYNDINGPNSDDSPVSSLSIGNE
YLDYTSNNSDEKEIVKLYDTHHEEFNKSSMPFVKSSNMIMNPNFNYEDNSS
FKGSLFETLWYFOQGVLLTSVYIFMLISIFDEIKFVFLTMMSSIKNKDHSD
TILQKQVLEYFVLPILSIIVTSIGFCEMIYIGTSAIKVHNNILYSILNAPLYIF
YNNGLIINREFIDISAPDYGFELKRIYKAFIIPRCILSILLIIMYRDICFIPFV
ILIIYFFVKFRCRCKEORLYLSCHTPLNCNIYSNLSGRNIINIYKKNYHLDYE
HYNNFRISYFFKWINIWASLYIKIFILLTTYIIMHPHYFASGIIKLYEKNVRI
LSTQYCSIFARGLVIIIFKLCDDYTHIEKEMCCVORLEEFPAKTSKNASNNKNEEL
LSTQYCSIFARGLVIIIFKLCDDYTHIEKEMCCVORLEEFPAKTSKNASNNKNEEL
NYITQYKERNENISDKISAIVEYKNVLSLSINSQDDESKKRYGKGFENVYSYK
KKIPLVNGTYKYIDEPKSLNIYALVKNOIGIVGKAGKSTLLSILGILINISOG
KITVGRDIRYNNRGEDSIIGILAQSSFFVYNNWIRTFIDPYNNTODEIVHALKLN
GINLKNLDLYKYMHPKQMSYKIIQTSKYVINGQSNNDTILLTNCIRKYSLSVRLYN
RHKYKILIDIEIPNLANSVHDELNSFLGKCSFNYYIIRNFPNNTVLIISHANT
LSCDDYIYVLRKGEITYRCSYEDVKQSELSHLEMDP"
23896..31533
/ gene="rRNA"
/ note="region containing small subunit, 5.8S and large
subunit rRNA genes and spacer regions"
23896..31533
/ gene="rRNA"
complement(join(31966..32476,32675..32775))
/ gene="MALIP3.04"
complement(join(31966..32476,32675..32775))
/ note="MALIP3.04, conserved hypothetical membrane protein,
len: 203 aa, similarity: P. falciparum chromosome 2,
PF0110W, 096126 predicted integral membrane protein (255
aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
in 191 aa overlap)"
/ codon_start=1
/ product="conserved hypothetical membrane protein,
MALIP3.04"
/ protein_id="CAB63559.1"
/ db_xref="GI:6594247"
/ translation="MKKSYTFINVTILLFLTLFLTYNYDYTFSKTFNNIKIDIN
REFKRIIAESEOKYPWEEDFCLILNEELIRPEHNDSPYLPHEHYENIDKINLSINS
TKIKWETIKKRQNYKEKETDNNHNRWDFMWHYKWANIYLYKYVHKLINITLKDLTNPI
HDKETITTWIKWIOEDIEYFLNQVWEVRLTLLELYFKNKE"
complement(32477..32486)
/ gene="MALIP3.04"
/ note="potential splice acceptor sequence"
complement(32669..32674)
/ gene="MALIP3.04"
/ note="potential splice donor sequence, aaa/gtatat"
join(36657..36743,36864..37343)
/ gene="MALIP3.05"
join(36657..36743,36864..37343)
/ gene="MALIP3.05"
/ note="MALIP3.05, hypothetical protein, len: 188 aa"
/ codon_start=1
/ product="hypothetical protein, MALIP3.05"
/ protein_id="CAB63560.1"
/ db_xref="GI:6594248"
/ translation="WRIKNSGIFIKILLICISFCVFCFNKCMISYRKDLWYSEN
CENYSIDSLARGSESKETVKVDIPNIELLKLNLNIEYEYKMKVIGSPMDNNLN
IANEVLKNIHSTENIENIFSLINDSSKSPVLTKLFKEGFIFFPHMLNNVPLKLLDLCQ
RNPLHIILGILVIAIYVFNFRNFC"
36744..36749
/ gene="MALIP3.05"
/ note="potential splice donor sequence, aag/gtatga"
36854..36863
/ gene="MALIP3.05"
/ note="potential splice acceptor sequence"
complement(join(38049..39995,40210..40284))
/ gene="garp"
complement(join(38049..39995,40210..40284))
/ gene="garp"
/ note="MALIP3.06, garp, len: 673 aa, similarity: almost

```

```

identical to GARP_PLAFF (678 aa), fasta scores: 97.6%
identity in 678 aa overlap"
/ codon_start=1
/ product="hypothetical garp protein"
/ protein_id="CAB63561.1"
/ db_xref="GI:6594249"
/ translation="MNVLFLSYNICILFFVVTCLNFSTKCFSGNLLKNQINILNKSPDS
ITGRLLNETELEKNDKSDSKSETLLKEKDEKDDVPTTSDNKLWANNNEISSSTDTP
TNIIVNDKDNENSVYDKKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
DKKKEENSEVSLYTKGOKPKNATEHGEENLYEEMVSEINNAOGGLLSLSPYOYRE
OGGCGIISVHETSNDTKDNKENISEDKKHQOEEMKLTLDKERQKQKEMKDEQE
KLEKKKQOEKKEKKQKKEKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
HKNKEETMQPDQTSSEETNNEIMVPLSPDVTTPPEHKEGHEKEHKEGHEKEG
BKHEEHEKEHKEHKEHKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
DKDGEIINLEDKACEQHITVESRPLSOPCKLIDPEOLTMDKSKVEKKNLSIQ
EOLICTIGRVNVVPRDNHKKMKAKIEAELOKQKHVDEKDKKESEKVESEKEVO
DDAEVEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
DDAEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
/ gene="garp"
complement(39996..40005)
/ note="potential splice acceptor sequence"
complement(40204..40209)
/ gene="garp"
/ note="potential splice donor sequence, aag/gtaaca"
join(45401..46396,46562..50233)
/ gene="MALIP3.07"

Query Match          9.4%; Score 88.6; DB 3; Length 67970;
Best Local Similarity 47.5%; Pred. No. 3.9e-05;
Matches 395; Conservative 0; Mismatches 424; Indels 12; Gaps 4;

Qy  41  taaattatgtgagagattcagtagaatctatataatacaacgcttactgatttgatt 100
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8558 TAAATATTAAATATATAAAATTAATAATATATATATATTAATAATAATAATATTATT 8617

Qy  101 tcaataattgcattgataatccaagtagagggtgatttaagaacattcttaacagaatatt 160
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8618 AAATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8677

Qy  161 cagttgtagataataaataaataatcttgcataagaagaaatatttggtttagcatcaa 220
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8678 TTAATAATATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8733

Qy  221 gtttgacaacggtgaaaaattcttaaggaggagaatatttttagaatggatgctgatg 280
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8734 AAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8793

Qy  281 atatttcataatccaagtagatttgataagcaaatctcgttttatggaggagaaattcattgg 340
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8794 GAATTAATAATCAATATATTATTAATAATAATAATAATAATAATAATAATAATAATA 8850

Qy  341 atttctcagcaactcaatagattgtagaccacaaagaaagaaatttagtatataaacaac 400
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8851 ATATTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8910

Qy  401 gagaaagtaataaaaataacttaactaatgatatacggaagatggttattgaatagatccta 460
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8911 AATATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8970

Qy  461 tacttgcacacccaacgctggcgtaaaaaagaaagaaagtttttcgataagtttaattgggata 520
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8971 TAATATATTAAACAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9026

Qy  521 gagatttagtacctgttgagattgatgttttgcaataagaggagcctcggtgatttca 580
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  9027 AATAATATTGCTTTATTATTATAATAATAATAATAATAATAATAATAATAATAATAATA 9086

Qy  581 aaatcggcttactcaataaagctactttttacagtagatttaaacgagaatgggaatcac 640
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  9087 TAATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9145

Qy  641 aaacccaataagtttaagcaaatatatttactcaagctattttacagaagattttttaaagaaa 700
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 9146 TAATAATTTAAATAATTAATAATTTAAATATATATAATAATAATAATAACGCTAATAA 9205

QY 701 aactcttattgatgatacacaaaataactacttctcaagagatgtgataaagaac 760

Db 9206 ATAATTTAAATATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9265

QY 761 gctatactcgcagagcgtctctaaatttttgagctaaatctacccctagttactta 820

Db 9266 AATATAATAAATTAATTAATTAATAATAATAATAATAATAATAATAATAATAATA 9325

QY 821 ttgaaaaactatatatttggtttattttacatttcaagtcctccctgggta 871

Db 9326 TAACATTTATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9376

RESULT 9

AC092304/c

LOCUS AC092304 Homo sapiens chromosome 19 clone LLNLFOS-20C5, WORKING DRAFT

DEFINITION SEQUENCE, 6 unordered pieces.

AC092304

AC092304.1 GI:14589493

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

KEYWORDS human.

SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 36977)

TITLE Sequencing of Human Chromosome 19

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 36977)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT -----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 31883, FOS36853

Center clone name: LLNL-FOS_20C5

Summary Statistics

Consensus quality: 31111 bases at least Q40

Consensus quality: 33034 bases at least Q30

Consensus quality: 33871 bases at least Q20

Estimated insert size: 42980; agarose-fp estimation

Estimated insert size: 36477; sum-of-contigs estimation

Quality coverage: 6.56 in Q20 bases; agarose-fp estimation

Quality coverage: 7.73 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will be preserved.

* 1 1580: contig of 1580 bp in length

* 1581 1680: gap of unknown length

* 1681 3748: contig of 2068 bp in length

* 3749 3848: gap of unknown length

* 3849 6070: contig of 2222 bp in length

* 6071 6170: gap of unknown length

* 6171 13635: contig of 7465 bp in length

* 13636 13735: gap of unknown length

* 13736 22951: contig of 9216 bp in length

* 22952 23051: gap of unknown length

* 23052 36977: contig of 13926 bp in length.

Location/Qualifiers

source 1. .36977

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/clone="LLNLFOS-20C5"

BASE COUNT 11122 a 7407 c 7254 g 10688 t 506 others

ORIGIN

Query Match 8.4%; Score 78.6; DB 2; Length 36977;

Best Local Similarity 46.7%; Pred. No. 0.0015;

Matches 433; Conservative 0; Mismatches 475; Indels 20; Gaps 5;

QY 27 atataatgagcctttaaattatgtagagagatcagtagaactctatactataaatacaaacgct 86

Db 34445 ATATATAATTTATATATAATAATAACATATATATATATATATATATATATATA 34386

QY 87 tactgatttgagtcbaattgctcattgataatcccaagtagaggtgatttaagcaatt 146

Db 34385 CATTATATATAATTTATATATAATAATAACATATATATATATATATATATATA 34326

QY 147 cttaacagaaatttcagttgtagataataagaataaaaaactcttgcctaa-----tgaaga 200

Db 34325 AATACATTTATATATAATTTATACATAATTTATAAATACATTTATGTATATATAACAATATATA 34266

QY 201 aaatattggttttagcatcaagttttgaacaaagcgtgaaaaatttctcaaggagaga-ata 259

Db 34265 CTATATAATTTATATATACTATATAATATATTTCTTATATAATTTATATATAATAAT 34206

QY 260 ttttgaagaaggatgctgatgatatttcoataatcccaagtagatttgataagcaattcggt 319

Db 34205 TATATGTTTATATAATTTATATATATATATATATATATATATATATATATATATAT 34146

QY 320 ttatggaggaaaatttcattggtattctcagcaactcctaataagaa-----ttgataagcc 373

Db 34145 ATATATAATATAATTTATATGTTTATATAATTTATATATATATAATAATAATAATTTATATGTTA 34086

QY 374 aaaaagaaaatttttagtataataaacaacgagaaagtaataataataataataataataatgata 433

Db 34085 TATAATTTATATATATAATAATAATTTATATGTTTATATAATAATAATAATAATAAT 34026

QY 434 tacggaagatgttattgaat----agatctatactctgcccccccaacggtgctgtaaaaa 489

Db 34025 ATATTTTATATGTTTATATAAATTTATATATAATAATAATAATAATAATAATAATAAT 33966

QY 490 aagaagtttccgataagtttaattgggataatagagatttagtacctgttgagattatgat 549

Db 33965 ATTATATATAATAATAATAATTTATATGTTTATATAATAATAATAATAATAATAATAAT 33906

QY 550 ttgcaataaagaggagctctgctggtatttcaaaatcggtctactcaataaagagctacttta 609

Db 33905 TATATTATATACTTACATATATATATATAATAATAATAATAATAATAATAATAATAATA 33846

QY 610 cagtagatttaacgagaaatggaatatacaacaacccaataagtttaagcaataattttac 669

Db 33845 TATTATATATAATAATAATAATTTGTTATATATATATATATATATATATATATATATAT 33786

QY 670 tcagctattttacaagatttttataaagaaaaactctatatattgatat---cacaaaaatt 726

Db 33785 TATAATATATACTTATATAGGTTTATGTTATATAATAATAATAATAATAATAATAATAAT 33726

QY 727 actaattactttcaagagatgtgataaagaagaacgctatactcagcagagagctctctctaaa 786

Db 33725 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 33666

QY 787 tattttgagctaaaatctacccctagtagtattactattgagaacactatactattgtttcat 846

Db 33665 TATAATAATTTATACAACTGTTATTAATAATAATAATAATAATAATAATAATAATAATACTGTA 33606

QY 847 ttactacttaagctcccttggttaggaggttataataaataaatgatatatttattttagta 906

Db 33605 TAATATACTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 33546

FEATURES


```
gene
CDS
/sub_species="O139 Bengal"
/db_xref="taxon:666"
190..1308
/gene="ORF41x9"
190..1308
/gene="ORF41x9"
/feature="putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase"
/protein_id="CAA69111.1"
/db_xref="GI:2244681"
/db_xref="SWISS-PROT:Q56598"
/translation="MKKALITGVTDGQSYLAPELLAKGYEVHGIKRRASSFNTORV
DHIVQDPHVDNASFLHYGDLTDSNLTIRILOQVKPDEVYNLGAQSHVAVFSPSEYIT
ADVDMGTLRLLEAIRLGLLEKTKFYQASTSELYGLVQETPOKETTPFYPRSPYAVA
KMAYIIVVNYRESYGYACNGILFNHESPRGETFVTRKITRGLANITAOLEKCLYM
GNMDLRMGHAKDHYRMQNMLOQDPDEFVIATGVQISVRFIENSAKELGVLTIF
EGQGVDEKGIIVTATEGDKAPALKVGDVVVQIDPRFRAVEETLLGDPSPAKQKLGWT
PEITVQENCAEMVEDLKVQARHALLKLHGHDPVPSVE"
1313..2239
/gene="ORF34x6"
1313..2239
/gene="ORF34x6"
/codon_start=1
/transl_table=11
/protein_id="CAA69112.1"
/db_xref="GI:2244682"
/db_xref="SPTREMBL:O34227"
/translation="MKILLTGARGVQGNILSLAGEYEHFELSPSSAQLNLLDSNAVH
DYIRHKPDWYIHAAGIVGGIOANNAQPVKFLVENMQMGLNLTVSNAQGVSQLNLS
SSCYPRDAHNPBELLKGELEPTNEGVALAKYSTRLCGEYISKENPSOIKYITIP
CNLYGRFPDNPISHMIPAVIRKTHEAKNLEIIDLWGGEARREPMYAEDEAEFV
FYADHKEMQNLNVLGTDYITNEYQAVASVIGYTGRRKHDLSKVPQGNRKLIDD
TKLAFGSHSTSLSEAGIEKTYQFYLYNGND"
2232..3404
/gene="ORF43x9"
2232..3404
/gene="ORF43x9"
/codon_start=1
/transl_table=11
/protein_id="CAA69113.1"
/db_xref="GI:2244683"
/db_xref="SPTREMBL:O34228"
/translation="MINYSLASSTWDQNEAIAQRIEVESDMFTMGKEVAQYKDFATF
FGAVALVSGSTANLLMIAALFTKNAKYKLERGDEVIVPAVSWSTTYPPLOQYGL
KVKFVIDRQTLNDLKDLEAATDNTRAILSVNLGNPNDFARMKDLIADRDLFILE
DNCSMGATNGKQAGTGVMGTFSSFSFHIAITWEGGCYVTDDEELKHILLIRAHG
WTRNLPKFNKTVGKSDDAFESEKFLVLPGLNVRPLEMSGALGIEQLKLPNFIEIRR
KNAELFQSLADHPFIEIOKETGESSWFGSLILKENAPYSRAELVKKLVNGIECRP
IWTGNFLKNKEVLEFFDEVAGTLEGAEYLDENGLFVGNHQNDIEAEIKLLAKVLV"
3404..3889
/gene="ORF18x8"
3404..3889
/gene="ORF18x8"
/codon_start=1
/transl_table=11
/protein_id="CAA69114.1"
/db_xref="GI:2244684"
/db_xref="SPTREMBL:O34229"
/translation="MYLEKEDFRGLVQHGVLFSIDLVLVDTONQLLGERVNRPAQGY
WFVPGGRYKNETLIKAFERICDELQOIFSYQDAKPLGLYDFHFNDSVFGDDISTHY
INAPHLIELSENTELTNLPSQHRHYRWVRIGLDAYDESVEHFSYKVFLELKLKRIAFTI
I"
3946..5382
/gene="ORF54x4"
3946..5382
/gene="ORF54x4"
/codon_start=1
/transl_table=11
/protein_id="CAA69115.1"
/db_xref="GI:2244685"
/db_xref="SPTREMBL:O34230"
```

```
gene
CDS
/translation="MIINAFYRDFLLSILSGKVQVFIGLTRLRLITEMLSEQVGIYI
LLTVVSLAFGFENPVGFGYGRYLITHMOKAELKTATNMLMLRGLTIPFALCLAIIV
FYMRFEQYFSIAEYSVFIVSVIALIHGVLNATNVLVSRVKFTYAVLTLFGLIL
SVYFQKQTAMANIYGLAFSQIFSFVILNITSEQESFVIKSAFNKMYKGVLP
FTLPVITLLQWQOTASRLIVEDISQAQLAFIAGVMTLGSALFSALESLSQFYM
PLYLKTIITNADOAMRTKAWNELADLLPEYIYAVLIYIATAPYLAKLVAEKFOQAYI
YAMIGAFIEFFRVITNLVYLSQSEVKTSTIMPLYLIGFMVMVIGLYIDVSRSPWL
PVILALSVLTLSIMFYNMKLLSIRFEFLNPAISLTLPLLVHLLSITPTLNSF
LSLLGGFYLLLIQYPFLKTKIKKMSAV"
5379..6395
/gene="ORF40x0"
5379..6395
/gene="ORF40x0"
/codon_start=1
/transl_table=11
/protein_id="CAA69116.1"
/db_xref="GI:2244686"
/db_xref="SPTREMBL:O34231"
/translation="MKKACLVVNDYMLCNKVFDEKIFRDNINIKYIKLKLERYGYD
LSTODINSIDSDVVIYAADIPKPLKRNEVKKSHLIWSESEFISPENYKVKHNSFN
KIFTWHDGLVDEKVKYFKINFHQFPNFINKNKKKLTCTLIAGNKKPHTSVGLYA
ERKALRMFKENHPSDFDLYGVGMDRYFSGSKIYRNLNRVPLLPKTKOKLTGRSYPS
YKGPIERKRPIMEKYKFSICYENVMDFCYITDKLDFDSFGCIPVYWGANNIETIIP
ENCFTDKRKFSLSYEDLYSYMKEMSDEEYIGYINNIQAFILKSDLSPFFKAETFAEKVVS
EVLN"
6395..7408
/gene="ORF40x1"
6395..7408
/gene="ORF40x1"
/codon_start=1
/transl_table=11
/protein_id="CAA69117.1"
/db_xref="GI:2244687"
/db_xref="SPTREMBL:O34232"
/translation="MKIRLIVAFKIEYGRKENGFPALYKAWFHGFEQJGYSIEGVFF
EDYDHKMLQTRIIIEKAMNSIDIVFIIQDQVDLDTLKYLSENNVTVPFGDDLWR
FDNWTSIYAPYSACITSDKFSIKKYEIGQGNIIYSQWASLEPKQKYSAPYKYDISF
IGSANRYKWFINLEKGIHVECFGYGKNGRVSYEDMELIFLQSKYNLINSISIF
DFRPMSSFSRFLSVKSYLYKDGKNSQTRARNEIPIVHGFOITDVPPTLEEVFKI
GKEIVCYNTIDEVEKLIRFYLINDYERECIKRAGTKRAREKHTYKARIKNFMVNDIKI
YDK"
7398..8855
/gene="ORF56x5"
7398..8855
/gene="ORF56x5"
/codon_start=1
/transl_table=11
/protein_id="CAA69118.1"
/db_xref="GI:2244688"
/db_xref="SPTREMBL:O34233"
/translation="MISKIGISKLTIEPHFFNELEPENYKVLTCSRSRLTWTFRDLAF
KLLYLEMDKNVSFSKEAYKEHIRAFSLGFGPKGOESKNSIEKRYDAFFTFNDISL
HGFATKSLIPLSHNGSTANGAHRVASAIILDKDVCYKLPVCDHLYDKFPYSRSVS
CDLDAIATKFEYADNVIYAFVWPTAGCFDEETIERIIPNIIYKRNIRMTPNGANLL
SOIYFGEPLWGTAEENNFRGSKNVTCTFDFMRVIAFRADSLDSVLQIKENIRQIF
NVGRHSIHITDTKGEAIRMARMIENDSIHFLNAYPNKYSTHEKLEAFKHHIDNFC
IGSDIILDSGMVLSIYGLREASDIDYLSIKSLEYKNEGLECHEKLEHDEEKNEL
IYNKYPFYENGLKFIATNOLYKNSRDEKVRDNCKMESLIENNQYKNIKAKLKQ
SIYVEKIKRLKIKTCLLASIGLYDLVKKIYKVLK"
8852..9865
/gene="ORF39x2"
8852..9865
/gene="ORF39x2"
/codon_start=1
/transl_table=11
/product="sugar transferase"
/protein_id="CAA69119.1"
/db_xref="GI:2244689"
/db_xref="SPTREMBL:O34234"
/translation="MSSPKISVMVSNGEKYLGFAIDSLKQTFSDFEFIINDGST
DKTLFRIKSNMKDDRIVLVSRENKGLIVSLNEGDLAKGOYIARMADDDISIKSRFE
KQIEFLDSDPDIQVCCGTWVEFGENIKSKKKMKMPQDDDLKAKLTFSVPFAHTVHIR
KKVIDKYINIDYLDGKYDAEDYKFWVDFSKYTLFSNVPEILLIRYRHQESISRAVDNKE"
```

NKERFRIISKIONEVLTSGVIVLTNEGKHNHFTLSLNERIINNVTDCDMIRAHLLKIS SSQIESQDSSAIERMLMLKYPFIYLILSTRDKOLSYLKIFDLMELKGAFLEKDRM EQP"		Wang, L., Huskic, S., Cisterne, A., Rothenmund, D. and Reeves, P. R. Direct Submission Submitted (18-DEC-2001) Microbiology, The University of Sydney, Sydney, NSW 2006, Australia	
gene	9868..10800 /gene="ORF35x9"	gene	/organism="Escherichia coli" /serotype="O55:H7" /db_xref="taxon:562"
CDS	9868..10800 /gene="ORF35x9" /codon_start=1 /transl_table=11 /protein_id="CAA69120.1" /db_xref="GI:2244690"	CDS	204..1199 /gene="gnc" 204..1199 /gene="gnc" /codon_start=1 /transl_table=11 /product="UDP-GlcNAC 4-epimerase Gne" /protein_id="AAL67550.1" /db_xref="GI:18266396"
gene	10797..11861 /gene="ORF41x8"	gene	/translation="MNDNVLLIGASGEVGTRLLETAADFNIKMLDQQSHFYPEITQ IGDVRDQALDQALAGDFTVLLAAAEHRDDVSPTSLYIDVNVQTRNVLAAMEKNGYK NIIFTSSVAVYGLNKNPDENHPDHPENHYGKSKMAQAEVLRMYNKAFTERSLTIIR PTVIFGERNEGVNLLKQJAGCKEMVAGATYKSNAYVGNIVFEFKYKLNVAAGY EYNYVDKPDLMNQLVAEVDQSLNKIPSMHLPYPLGMLGCGYCFDLSKITGKKYAV SSVRVKFCATTQFDATKVVHSSGFVADYITLSQLDRTLQYEFVHAKKDDITTFVSE"
CDS	10797..11861 /gene="ORF41x8" /codon_start=1 /transl_table=11 /product="Rfc-like protein"	CDS	1442..2335 /gene="galf" 1442..2335 /gene="galf" /codon_start=1 /transl_table=11 /product="UDP-glucose pyrophosphorylase Galf" /protein_id="AAL67551.1" /db_xref="GI:18266397"
Query Match	8.2%; Score 77.2; DB 1; Length 16265;	gene	/translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIDVDPKPMIOYIVDEI VAAGIKELLVTHASKNAVENHEDTSYELESLEQVRKQLLAEVQSCIPGVTIMVY QVLEGLGHSILCARPAIGNPVPVLPDVIDDASADPLRVNLAAMIARFNETGRS QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPOTLDSIDMAVGRYLS ADTWPELERTOPGAWGRIQLTDAELAKKOSVDAMLMTGDSYDCGKKMGVMAQAFVY GLNLREGAKFLKNIQLAK"
Best Local Similarity	57.2%; Pred. No. 0.003;	CDS	2716..3810 /gene="wbgm" 2716..3810 /gene="wbgm" /codon_start=1 /transl_table=11 /product="putative galactosyltransferase WbgM" /protein_id="AAL67552.1"
Matches 183; Conservative	0; Mismatches 128; Indels 9; Gaps 2;	gene	/translation="MVKILHVHLSGDISGAQVSLDEMHTLSQEFQOQSMVCSKBGRLA EQARCFQVCTHIPTLTREISLFKDCASLFQLYKIKKEKFDIVHTSSKTGFLGRVA AKLAGTKKIVHTVHGFAFPSTENKLIKFIYFLMELIASYCSNIIIVNSEDERTARKY FVKNKSKLLIINNAIDVKYKNDKDKDLPFKIVMGRLCDOKNPLLLIEAKDL ESNIHYDIDGGLPKVKLLEKINOYNIADKVSFLGWDIDAVEEHLKYKDLFPVLSRWEG MPLAMLEMAAKVPVLSSDIEANKYLIETKTAGVVVFKDEDSKDLARKKINVLHNPRLRN NLAHKAYQALIEDFDLTKRTKILLESYLN"
QY 17	ttatgtcggtatataatgagcccttaaatattgtgagagattcagtagaattctatattaa 76	gene	3831..4949 /gene="gmd" 3831..4949 /gene="gmd" /codon_start=1 /transl_table=11 /product="GDP-mannose dehydratase Gmd" /protein_id="AAL67553.1" /db_xref="GI:18266399"
Db 8877	TTATGTCTGTTTATAATG---GTGAAAAATATCTTGCAGAGCCATAGATAGTATTCTTA 8933	CDS	/translation="MTKVALITGVTDGQSGYLAEFLEKGYEYHGIKRRASSFNTERV DHIYQDPHSAPNPHLDLTDTSNITRIQLEVKVPDEYVNLGAMSHVAFVSFPEYT ADVDATGLLEAIRGLKKTREYQASTSELYGLVQIPQKETTTPFPSPRYAVA KLAYWITVNYRESYGIYACNGILFNHESPRGETFVTRKTIATIAIQAQLESCLYL GNMDSLDRMGHAKDYVRMOMMLQEQEPDFVATGVOYSVREFVENTAEOGLGKLSF EGKVDEKGIYVSVTGDKAPGVGTGVIVADPRYFRPAEVTLLGDPKAKAHEKLGHK PEITLREMISEMVANDLQTKAKKVVLLKSHGFNANLAQE"
QY 77	atcaaacgcttactgatttgatttcataattgtcattgataatcccaagtagagtgatt 136	gene	5588..7045 /gene="manC"
Db 8934	AGCAACATTTTCTGATTTTGAATTTTATTTATTCATTAAATGACGGTTCGAC-----TGATA 8987		
QY 137	taaaagcaattctaacagaatttcagttgtagataatagaataaaaaattctgcttaag 196		
Db 8988	AGACACTAGAGATANTAAAGTCATATATGAAGAAGGATGACCGTATTGTATTGTTTCAA 9047		
QY 197	aagaaaatttggtttagtcacagtttgaaacaaacggtgaaattcttaaggagagaat 256		
Db 9048	GAGAAACAAAGGCGTAATAGTAAGCCTAAATGAAGCGCTAGATTAGCTAAAGGCGCAAT 9107		
QY 257	atatatttagaagtagctgtagatatttcataatcccaagtagattgataagcaaatcc 316		
Db 9108	ACATTGCTAGAAATGGATGGCAGATGATATTTCAATAAAGTCACGTTTTTGAAGCAAAATTG 9167		
QY 317	gtttatggaggaaaattca 336		
Db 9168	AGTTTTTAGATAGTAATCCA 9187		
RESULT 12			
AF461121	27730 bp DNA linear BCT 23-JAN-2002		
LOCUS	Escherichia coli serotype O55:H7 O-antigen gene cluster, complete		
DEFINITION	sequence.		
ACCESSION	AF461121		
VERSION	AF461121.1		
KEYWORDS	GI:18266395		
SOURCE	Escherichia coli.		
ORGANISM	Escherichia coli.		
REFERENCE	1 (bases 1 to 27730)		
AUTHORS	Wang, L., Huskic, S., Cisterne, A., Rothenmund, D. and Reeves, P. R.		
TITLE	The O-antigen gene cluster of Escherichia coli O55:H7 and		
JOURNAL	Identification of a new UDP-GlcNAC epimerase gene		
REFERENCE	2 (bases 1 to 27730)		

AUTHORS TITLE JOURNAL FEATURES source		Wang, L., Huskic, S., Cisterne, A., Rothenmund, D. and Reeves, P. R. Direct Submission Submitted (18-DEC-2001) Microbiology, The University of Sydney, Sydney, NSW 2006, Australia	
gene	10797..11861 /gene="ORF35x9"	gene	/organism="Escherichia coli" /serotype="O55:H7" /db_xref="taxon:562"
CDS	10797..11861 /gene="ORF35x9" /codon_start=1 /transl_table=11 /protein_id="CAA69120.1" /db_xref="GI:2244690"	CDS	204..1199 /gene="gnc" 204..1199 /gene="gnc" /codon_start=1 /transl_table=11 /product="UDP-GlcNAC 4-epimerase Gne" /protein_id="AAL67550.1" /db_xref="GI:18266396"
gene	10797..11861 /gene="ORF41x8"	gene	/translation="MNDNVLLIGASGEVGTRLLETAADFNIKMLDQQSHFYPEITQ IGDVRDQALDQALAGDFTVLLAAAEHRDDVSPTSLYIDVNVQTRNVLAAMEKNGYK NIIFTSSVAVYGLNKNPDENHPDHPENHYGKSKMAQAEVLRMYNKAFTERSLTIIR PTVIFGERNEGVNLLKQJAGCKEMVAGATYKSNAYVGNIVFEFKYKLNVAAGY EYNYVDKPDLMNQLVAEVDQSLNKIPSMHLPYPLGMLGCGYCFDLSKITGKKYAV SSVRVKFCATTQFDATKVVHSSGFVADYITLSQLDRTLQYEFVHAKKDDITTFVSE"
CDS	10797..11861 /gene="ORF41x8" /codon_start=1 /transl_table=11 /product="Rfc-like protein"	CDS	1442..2335 /gene="galf" 1442..2335 /gene="galf" /codon_start=1 /transl_table=11 /product="UDP-glucose pyrophosphorylase Galf" /protein_id="AAL67551.1" /db_xref="GI:18266397"
gene	10797..11861 /gene="ORF41x8"	gene	/translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIDVDPKPMIOYIVDEI VAAGIKELLVTHASKNAVENHEDTSYELESLEQVRKQLLAEVQSCIPGVTIMVY QVLEGLGHSILCARPAIGNPVPVLPDVIDDASADPLRVNLAAMIARFNETGRS QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPOTLDSIDMAVGRYLS ADTWPELERTOPGAWGRIQLTDAELAKKOSVDAMLMTGDSYDCGKKMGVMAQAFVY GLNLREGAKFLKNIQLAK"
CDS	10797..11861 /gene="ORF41x8" /codon_start=1 /transl_table=11 /product="Rfc-like protein"	CDS	2716..3810 /gene="wbgm" 2716..3810 /gene="wbgm" /codon_start=1 /transl_table=11 /product="putative galactosyltransferase WbgM" /protein_id="AAL67552.1"
gene	10797..11861 /gene="ORF41x8"	gene	/translation="MVKILHVHLSGDISGAQVSLDEMHTLSQEFQOQSMVCSKBGRLA EQARCFQVCTHIPTLTREISLFKDCASLFQLYKIKKEKFDIVHTSSKTGFLGRVA AKLAGTKKIVHTVHGFAFPSTENKLIKFIYFLMELIASYCSNIIIVNSEDERTARKY FVKNKSKLLIINNAIDVKYKNDKDKDLPFKIVMGRLCDOKNPLLLIEAKDL ESNIHYDIDGGLPKVKLLEKINOYNIADKVSFLGWDIDAVEEHLKYKDLFPVLSRWEG MPLAMLEMAAKVPVLSSDIEANKYLIETKTAGVVVFKDEDSKDLARKKINVLHNPRLRN NLAHKAYQALIEDFDLTKRTKILLESYLN"
CDS	10797..11861 /gene="ORF41x8" /codon_start=1 /transl_table=11 /product="Rfc-like protein"	CDS	3831..4949 /gene="gmd" 3831..4949 /gene="gmd" /codon_start=1 /transl_table=11 /product="GDP-mannose dehydratase Gmd" /protein_id="AAL67553.1" /db_xref="GI:18266399"
gene	10797..11861 /gene="ORF41x8"	gene	/translation="MTKVALITGVTDGQSGYLAEFLEKGYEYHGIKRRASSFNTERV DHIYQDPHSAPNPHLDLTDTSNITRIQLEVKVPDEYVNLGAMSHVAFVSFPEYT ADVDATGLLEAIRGLKKTREYQASTSELYGLVQIPQKETTTPFPSPRYAVA KLAYWITVNYRESYGIYACNGILFNHESPRGETFVTRKTIATIAIQAQLESCLYL GNMDSLDRMGHAKDYVRMOMMLQEQEPDFVATGVOYSVREFVENTAEOGLGKLSF EGKVDEKGIYVSVTGDKAPGVGTGVIVADPRYFRPAEVTLLGDPKAKAHEKLGHK PEITLREMISEMVANDLQTKAKKVVLLKSHGFNANLAQE"
CDS	10797..11861 /gene="ORF41x8" /codon_start=1 /transl_table=11 /product="Rfc-like protein"	CDS	5588..7045 /gene="manC"


```

Db 12172 TCAAAATTACCGGAATATTATTCTATTACAGGTTGCATGA 12213

RESULT 13
AB012956
LOCUS AB012956 46721 bp DNA linear BCT 16-OCT-1999
DEFINITION Vibrio cholerae genes for O-antigen synthesis, strain MO45,
complete cds.
ACCESSION AB012956
VERSION AB012956.1 GI:3724295
KEYWORDS
SOURCE Vibrio cholerae (strain:MO45) DNA.
ORGANISM Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (sites)
AUTHORS Yamasaki,S., Shimizu,T., Hoshino,K., Ho,S.T., Shimada,T., Nair,G.B.
and Takeda,Y.
TITLE The genes responsible for O-antigen synthesis of vibrio cholerae
O139 are closely related to those of vibrio cholerae O22
JOURNAL Gene 237 (2), 321-332 (1999)
MEDLINE 99453293
REFERENCE 2 (bases 1 to 46721)
AUTHORS Yamasaki,S., Shimizu,T. and Hoshino,K.
TITLE Direct Submission
JOURNAL Submitted (14-APR-1998) Shinji Yamasaki, International Medical
Center of Japan, Research Institute: Toyama 1-21-1, Shinjuku, Tokyo
162-8655, Japan (E-mail:shinji@i.imcj.go.jp, Tel: +81-3-3202-7181,
Fax: +81-3-3202-7364)
COMMENT Sequence updated (24-Sep-1998).
FEATURES
source
    1..46721
        /organism="Vibrio cholerae"
        /strain="MO45"
        /db_xref="taxon:666"
        /complement(7866..8810)
        /gene="gmhD"
        /complement(7866..8810)
        /gene="gmhD"
        /note="probable LPS core synthesis"
        /codon_start=1
        /transl_table=11
        /protein_id="BAA33585.1"
        /db_xref="GI:3724296"
        /translation="MTIVTGAGMIGSNIIKALNERGITDILVVDHLKNGRKFKNLVD
        LQADYMDRDFLAQIMAGDDFGSIDAIFHEGACSATTEWDGYVMLNNYYSKEILLH
        YCLDREIPLYASSAATYGETDFEIEPOYEGALNVGYSKQOFDNYVRLWLDKQIH
        DETLSOITGERFVNYGPREQHKGSMAFHLNOMNAGENKFLFAGSENFRKDFVY
        GDVAANLWFLDHGVSGIFNCGTGKAESEFNEVAKAVIAPHGRGEVETIPFPDHLKGA
        YQEFTEADUTKLRAAGCDYQFKSVAEGVAEYMALINRK"
        /complement(8924..10441)
        /gene="wbfa"
        /complement(8924..10441)
        /gene="wbfa"
        /note="unknown"
        /codon_start=1
        /transl_table=11
        /protein_id="BAA33586.1"
        /db_xref="GI:3724297"
        /translation="MHKPTIISVIALTLCCGGESGNSCNTTPPVKYFNVSFLDLDN
        IAINOTNCOIFYNEDQTKIVAYSAPENNTOIVVHDEHGNFVRYSNSTETNGLS
        STFRQSDVPQNGYVSFVYSEGRDNFVTSKDFIPESFSIYAKNNKNNNSCLSIGS
        GNNNSPOTSDYKSIHIIPDENKLFFFSLNHNQSLSEPTNKLGRDYSVNSRRGRPLL
        TAYERTNQDYLKLFKFRATNALGSGQIEDLIDIDQDRDTPWTPSDVELSSAQLF
        VDGRQNPAPYLMQSLNLTSSDEKYSYSLNLDVNDYLYLKLKREISPPSPYWKFOHV
        SRTGSSTSHLSAENLIDGFPNEPPTGLICNDHNTQCILKITHQKTQDNTLRLLLID
        SEUTVSGGPIFTKQIYISPHNELPMKFPDPSILDQELKEIKDFSVSWIKTFTSSSIT
        ETELYQDLYKRTLSGLSDPSIDNIPLLKNLSAQDQDQDLKRPYTWLLEE"
        /complement(10522..12714)
        /gene="wbfb"
        /complement(10522..12714)
        /gene="wbfb"
        /note="unknown"
        /codon_start=1

gene
    complement(12711..13577)
CDS
    complement(12711..13577)
    /gene="wbfc"
    /complement(12711..13577)
    /gene="wbfc"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA33588.1"
    /db_xref="GI:3724299"
    /translation="MCQTDLLRLRSPFKICRSPIKKRAMSFITSRGWLTIAAAISGH
    SLCAAPLATSLVVRVAAHNOAIIHSTRLSFSEPMRLDSAVMOTLEOAGLNTQOQWEP
    AGFLDLSHAFLEKRDVLLKLADQOASSAPPTQOALWASLIAQLROAEFAKRLFISVDP
    WTRIAPHQHNRLNGSWLLTLNKSSTQSVYGVAVNQPGDYVHNRLSADYAHAGLLD
    EQISEIVVIQPDGIAQKHAAYWNNQDFNEVAPGAIVTVPLPKRAFFDPTVTDADLQ
    LVIELLRNRLPL"
    /complement(13577..14266)
    /gene="wbfd"
    /complement(13577..14266)
    /gene="wbfd"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA33589.1"
    /db_xref="GI:3724300"
    /translation="MHRLYLAFQCLFLMCCSNSSQHLSETMQLAIFGTAPVPISPQ
    KVAGPIYASLYIQONDPLMPLAWAEPTNRNHRPALKWWVSAGOEMLVTEGGRIKT
    LNLGDANLISLESDDTPLLGLHRSTPREMHFYLSWQPGYHFRYLAHSQFITLGEQ
    KQLQPNGESRTLLRIDRQITIAALNQYHNOYWLPESTPSTGAVIASQQALAPGTHRYALA
    VAKSYLNOGGI"
    /complement(14638..14638)
    /gene="orif6"
    /complement(14638..14638)
    /gene="orif6"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA33590.1"
    /db_xref="GI:3724301"
    /translation="MATATAAAVATATAVVVAVPTAAVLAIVEPPALAAVSSVAVAKPL
    TSSALNTSAEFFIVISMKDIIYTDVHSPRETRTCKR"
    /complement(14935..15252)
    /gene="wbfe"
    /complement(14935..15252)
    /gene="wbfe"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="BAA33591.1"
    /db_xref="GI:3724302"
    /translation="MYOSIHASLARNVFKTSPKLFSTPLCQCQVALMTSSLTSSLLV
    PSVAEPIYAGISLAYTNAEYOPNHAASFTDGSPIAVALQGGYFFSDYLGLEARYSTGV
    QRSSGVNIDNLSSAFVKLNLPMSAQVSLYGLAGYSYFSLDRQGVGSVDESFGSSGIGI
    HYALDSRAFAFDFFINHANGDEARLNTVNLFSQLKF"
    /complement(15645..18380)
    /gene="wbfb"
    /complement(15645..18380)
    /gene="wbfb"

```


Db	68007	AAATATCTAAATATCGATATTTATATATATATAAT	67973
RESULT	15		
AX344555/c			
LOCUS	AX344555	349980 bp	DNA
DEFINITION	Sequence 6 from Patent WO0200932.		linear
ACCESSION	AX344555		
VERSION	AX344555.1	GI:18492441	
KEYWORDS			
SOURCE	synthetic construct.		
ORGANISM	artificial sequence.		
REFERENCE	1 (sites)		
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.		
TITLE	Diagnosis of Known genetic parameters within the mhc		
JOURNAL	Patent: WO 0200932-A 6 03-JAN-2002;		
	Epigenomics AG (DE)		
FEATURES	Location/Qualifiers		
source	1. .349980		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	/note="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>-split as follows.-seq 01 0.000.001 to 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length of seq 2: 3.673778 <223>-split as follows.-seq 14 0.000.001 to 0.349.980-seq 15 0.300.001 649.980-seq 16 600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001 1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001 2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001 2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001 3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001 3.673.778"		
BASE COUNT	95599 a	4385 c	78257 g 171739 t
ORIGIN			
Query Match	8.1%;	Score 76.2;	DB 6; Length 349980;
Best Local Similarity	44.6%;	Pred. No. 0.002;	
Matches	341;	Conservative 0;	Mismatches 423; Indels 1; Gaps 1;
QY	74	taaatcaaacgcttactgattgttgagtttcataaattgctcattgataatccaagtagaggtg	133
Db	215364	TAAATATACATATAAAATATATATATAAAATATATAAAATATATAAAATATATAAAAT	215305
QY	134	atttaagaacaattcttaacagaatttcagttgtagataataagaataaaaacttgcgcta	193
Db	215304	ATATATATAAAATATATATAAAATATATAAAATATATAAAATATATAAAATATATATAT	215245
QY	194	atgaagaaaattgtgttcagcatcaagtttgacaacgcggtgaaaatttcttaagggag	253
Db	215244	AAATATATATATAAAATATATATAAAATATATATAAAATATATAAAATATATAAAATA	215185
QY	254	aatatattttgaagtggtgctgatgatatttcataatccaagtgagtttgataagcaaa	313
Db	215184	TATATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAATA	215125
QY	314	ttcggttttatggaggaaaattccttggtatttcttcagcaactcctaagaattgatagacc	373
Db	215124	TAAATATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATA-T	215066
QY	374	aaaaaggaatttagtatataaacacgcgagaaagtaataaaaatacttaactaatgata	433
Db	215065	AAATATATATATATATAAAATATATAAAATATATAAAATATATAAAATATATAATAA	215006
QY	434	tacggaagatggttattgaatagatctctacttgccaccacgcgtgggtcgctaaaaaaga	493

	D	b	215005	AATATATATATATAAAATATATATATATATATAAATATATATAAAAATATATATAAC	214946
	Qy	494	aagtttccgataagttaattgggatagagaatttagtacctgttggaagtattatgatittcg	553	
	D	b	214945	AAAAATATATATACA AAAATATATATATAA AATATATATAA AA ATATATATAA TATATA	214886
	Qy	554	caataaggaggagcctctgcgtgtattccaatatoggcttactccaataagaacctttaccagt	613	
	D	b	214885	TATATATAAAATATATATAA AATATATATACATATAA AATATATATATACTATATATAA AAT	214826
	Qy	614	atagattaaacgagaatggaatacacacaaccaaatgaagtttaaagccaatatatttactcaag	673	
	D	b	214825	ATAAAATAATATATATAA AAATATATATATAA AATATATATAA AATATATATAA ATA	214766
	Qy	674	ctatttcacaagattttctaagaagaaaatccttatgtgatccacaaaaattactaat	733	
	D	b	214765	TACATATAAAATATATATAA AATATACATATAA AATATACATATAA AATATACATATA	214706
	Qy	734	accttcraagatgtgataaaagaaacgcctactacgacgaagcgtctctcaaatttttg	793	
	D	b	214705	AAATATACATATAA AATATACATATAA AATATATATAA AATATATATAA AATATATATAT	214646
	Qy	794	agctaaatcatcccctagttactacttaga aaacatatattt	838	
	D	b	214645	ATAAAATATATATATAA AATATAA AATATATATAA AATATAA AATATATAA ATATATAT	214601

Search completed: August 2, 2002, 13:35:08
Job time: 10289 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 10:39:54 ; Search time 1628.88 Seconds
(without alignments)
7780.586 Million cell updates/sec

Title: US-09-900-038A-2
Perfect score: 939
Sequence: 1 atgaattatgatcattat.....gaggagagaacaagtgc 939

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89.8	9.6	1101	12	CNS00EVL
c 2	87.2	9.3	1101	12	CNS0039G
c 3	77	8.2	1101	12	CNS00EVL
c 4	71.6	7.6	1248	12	B11336
5	69.6	7.4	893	12	CNS013XE
c 6	68.8	7.3	928	12	CNS0102F
7	68.4	7.3	928	12	CNS00DKY
8	68.4	7.3	1101	12	CNS0039G
c 9	68.2	7.3	1101	12	CNS017KE
10	67.2	7.2	1001	12	CNS01400
c 11	67.2	7.2	1101	12	CNS00EVL
c 12	66.8	7.1	1025	12	CNS014JZ
c 13	66.2	7.1	1169	12	CNS06KHQ
c 14	65.4	7.0	945	10	BM358135
15	65.4	7.0	1092	12	CNS020K7
16	64.2	6.8	910	12	CNS01G8P
17	63.8	6.8	1101	12	CNS00EPO

18	63.6	6.8	987	12	CNS014PQ
c 19	63.4	6.8	1101	12	CNS001FB
c 20	63	6.7	782	9	BB622182
c 21	62.2	6.6	867	12	CNS075BG
22	62	6.6	1101	12	CNS0021J
c 23	61.8	6.6	756	12	BH575203
24	61.6	6.6	777	12	CNS025WB
25	61.6	6.6	916	12	B09084
26	61.6	6.6	945	12	CNS04DOK
27	61.6	6.6	994	12	CNS04NOJ
28	61.4	6.5	1085	12	CNS0161R
c 29	61.2	6.5	938	12	CNS006TJ
c 30	61.2	6.5	1101	12	CNS0175J
31	60.8	6.5	1101	12	CNS008X3
32	60.8	6.5	1101	12	CNS00YWL
c 33	60.8	6.5	1101	12	CNS05AB2
34	60.8	6.5	1200	12	CNS016CO
35	60.6	6.5	1190	12	CNS02ON7
36	60.4	6.4	734	12	CNS01OMP
37	60.4	6.4	849	9	AL578920
c 38	60.4	6.4	1101	12	CNS014XJ
39	60.2	6.4	886	12	BH177277
40	60.2	6.4	886	12	CNS07JUX
41	60.2	6.4	996	12	CNS00FUH
42	60.2	6.4	1101	12	CNS003BD
43	60.2	6.4	1101	12	CNS00EO7
c 44	60.2	6.4	1190	12	CNS02ON7
c 45	60	6.4	1101	12	CNS00BO1

ALIGNMENTS

RESULT 1

CNS00EVL

LOCUS

DEFINITION

CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

AL069706.1 GI:4949849

VERSION

GSS.

KEYWORDS

fruit fly.

SOURCE

Drosophila melanogaster

ORGANISM

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

REFERENCE

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

AUTHORS

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr

TITLE

Direct Submission

JOURNAL

- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1..1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

```
/clone="BACR29B23"
/Note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 9.6%; Score 89.8; DB 12; Length 1101;
Best Local Similarity 34.0%; Pred. No. 1.7e-05;
Matches 214; Conservative 104; Mismatches 311; Indels 0; Gaps 0;

QY 138 aaagcaattcttaacagaatattcagttgttagataatagaataaaaaattcttcttaataga 197
DB 467 MMMWAAATTTWAAAWAAWAAATATTAATWAAATAAAWAAWAAATTTTMMWWTWATTTT 526
QY 198 agaaatatgttttagcatcaagtttgacaaagcggtgaaattcttaagggagaata 257
DB 527 WAWWTWATAAAAAAAWAAATATTTAAWAAWAAATWAAATTTWAAWWTATAT 586
QY 258 tatttttagaattgagctgtagatattctcatatccaaagttagattgataagcaaatcg 317
DB 587 AATWATAAATWATAATATAAAAAAATATTTTWTATAAAATTTTAAATAATTTAAT 646
QY 318 ttttttagggagaaaattcattggatttctcagcaactcctaataagattgataagcaaaa 377
DB 647 TAWTTATAAATAAAWTAATTTWMTAAATTAATAATTTWAAATAWAAAAAATAAAW 706
QY 378 aggaatttttagtataacacagagaaagtaataataatacttaacttaagcatatagc 437
DB 707 ATWAAWAAATWATAWTAATTTAAWAAWAAATAAAWAAWAAATWAAWATAATWATATA 766
QY 438 gaagatggttaagatagatctatctatcctgccaccacgctggtgcgtgtaaaagaagt 497
DB 767 WTTTAAWWTAAWWTATATAWATAWAAATAWAAWAAWATAATAWATAWATAWAAWAA 826
QY 498 tttagaattgaattgaattgagatttagtacctgttggaattgattttccaat 557
DB 827 WATAWWTATATAWAAATAWAAWAAATAATTTAAATWATAWATAWAAWAAATAWAAWTT 886
QY 558 aagagagctcgtgctgattccaaatcgcttactcaataaagttactttcagctatag 617
DB 887 WMTTWTWAAWATATAAAWATAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWTT 946
QY 618 attaacgagaaatggaatatcacacacacaaatgaattgaagcaatatattactcagctat 677
DB 947 TTATTAATWTTATWTTWATTTWATTTWATTTWATTTWATTTWATTTWATTTWATTT 1006
QY 678 tttaagaatttttaaaagaaaatcttattgattgataacacaaaattactaattactt 737
DB 1007 ATTATTAWWTATATTTTAAWAAWTAATATATWATAWWTATATAWAAWAAATTAWTT 1066
QY 738 tcaagagatgtgataaagaaacgctata 766
DB 1067 ATATATTAATAWAAWAAWAAWAAWAAWAA 1095

RESULT 2
CNS0039G/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
DIRECT SUBMISSION
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL
```

```

Db      475 TTTTWTW 470

RESULT      3
CNS000EVL/c
LOCUS
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
              BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
AUTHORS     Direct Submission
TITLE       Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL     BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            Pl and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29B23"
                     /note="end : T7"
     BASE COUNT       419 a    91 c    60 g    299 t    232 others
     ORIGIN

Query Match      8.2%  Score 77;  DB 12;  Length 1101;
Best Local Similarity 33.7%  Pred. No. 0.0021;
Matches 212;  Conservative 105;  Mismatches 310;  Indels 2;  Gaps 1;

QY 281 atattccatccaaagtagatttgataagcaaaattcgcttttatgaggaaaaattcattgg 340
      :||| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1094 WTWTWTWTWTWTWTATWTATATATATATAAATAATTTWTTWTATATWTATWTATWTATATA 1035

QY 341 attctcagcaacctctaatagaattgatagaccacaaagaaatttagtatataaacaaac 400
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1034 TTATWTWTWTAATAATATATWTTTAAATATAATATATATATATATATATATATATATATATA 975

QY 401 gagaagtaacaaatactactaactaatgatatacgaagaattatttgatagatcata 460
      : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 974 AWTWAAATWAAATWATAWAWTTTAAATAAAATAATATATATATATATATATATATATATAT 915

QY 461 tacttgcaccacccaacgctggcgctgaacaaagatttttcogataagttaatgggatata 520
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 914 TWTATWTWTWTATATATWTWAAAAAAWAAWWTATATTTTWTTA--TWTATWTATATATW 857

QY 521 gagatttagtacctcgttgagaattatgattcttgcacaaagaagagctcgtcgattcca 580
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 856 ATTTTWTWTWTATWTATATATATATATATATATATATATATATATATATATATATATATAT 797

```

[illegible]

Matches 233; Conservative 0; Mismatches 289; Indels 0; Gaps 0;				
Qy	340	gatttctcagcaacttaataagaattgatagaccacaaaggaatttagtatataacaa	399	
Db	716	GATTTTAAAAAATAAATAAAGATATTAATAAAAAAATTAATAAAAAATNAAA	657	
Qy	400	cgagaaagtaataataacttaactaatacgaatacgcgaagatgtttagatgatct	459	
Db	656	ATAAATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	597	
Qy	460	atacttgccacccacgcggtgcgtataaagaaagtttctcgataaagtttaaggat	519	
Db	596	ATATATTATAAAAAATAAAAAATAAATAAATAAATAAATAAATAAATAAATAA	537	
Qy	520	agagatttagtcctggtgaagatcatgtttgcgaataagaggcgtcgtgatttc	579	
Db	536	AAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	477	
Qy	580	aaatcggttactcaataaagtaactttcacagtagattaaacgagaatggaatca	639	
Db	476	NNTTTTNTAAAAAATAAAAAATAAATAAATAAATAAATAAATAAATAAATAA	417	
Qy	640	caaaccaataaagttaagcaatatttactcagctattttcaagatttttataagaa	699	
Db	416	ATATTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	357	
Qy	700	aaatcttatatgatcatcaaaaattactaattacttcaagagtagtgcgataagaa	759	
Db	356	ATAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	297	
Qy	760	cgctatactcagcagagcgtcctcaataatttttgagctaaatctaccocctagt	819	
Db	296	AANNATANTTATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	237	
Qy	820	attagaaaaactatatttggtttatatttatacttcttaagtcct	861	
Db	236	TTTTAAATTAATTTTAAATTAATTTTAAATTAATTTTAAATTAATTTTAA	195	

RESULT 5
LOCUS CNS013XE 893 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN10102 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL103436
VERSION AL103436.1 GI:5615047
KEYWORDS GSS:
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 893)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES
source Location/Qualifiers
1..893
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"

/clone="BACN10102" /note="end : SP6"				
BASE COUNT	372 a	21 c	33 g	308 t 159 others
ORIGIN				
Query Match	7.4%; Score 69.6; DB 12; Length 893;			
Best Local Similarity	36.5%; Pred. No. 0.034;			
Matches 295; Conservative 84; Mismatches 426; Indels 4; Gaps 2;				
Qy	127	agagggtatttaagaacaattcttaacagaataatcagttgtgagataataagaataaaaaatc	186	
Db	23	AAAAATNTTCTCTATTATTHMTATWMMWTATMTTCTTTTMMMTAATAATTATTAATTTT	82	
Qy	187	tltgcttaatgaagaaataattggttttagcatcaagtttgaacaaagcgtgaaaaattct	246	
Db	83	TATATATATWTATTATATAMAMATATAAAMATATAAAAAAAMMMAMATTTATAAAAAAT	142	
Qy	247	aaggggagaaatattttttagaagatgcctgatgatatttccatataccagtagatttgat	306	
Db	143	TWAMTTAAATTWATTTATATAAAATTAMATTTATAMAAATAAATTTATATAAAAAA	202	
Qy	307	aagcaaaatcgttttatggaggaaatattcattggatttctcagcaactctaataagaattg	366	
Db	203	AAAAAAMAMTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	262	
Qy	367	atagaccaaaggaatttagtatataaacaacagagaagtaataaataatacttaact	426	
Db	263	CTAAAAATTTATMTAMAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	322	
Qy	427	aatgatatacggagaatgttattgaatagatctatacttgcgcacccacacgcgtggcgta	486	
Db	323	MAAMATANTNTAAAAATATNTTNAMTAAATAAATAAATAAATAAATAAATAAATAAATAA	379	
Qy	487	aaaaagaaagtttctcgataagtttaatgggatatagagatttagtacctgttgaagattat	546	
Db	380	MTAATTAAMTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	439	
Qy	547	gatttgcgaataagaggagcgtcctgctgatttcacaaatcggtctactcaataaagactt	606	
Db	440	ATWTTTTTTCATANCACCTCCTATTTCATNAACACCACCAATTTTCCTCTATTATTTT	499	
Qy	607	ttacagtatagattaaacagagatggaatatcacacaaccaataagtttaagcaatatatt	666	
Db	500	TTWATCTATATTTTMAAAATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAA	559	
Qy	667	tactcagctattttcaagaatttttataaagaataattcttatatgatatcacaaaaatt	726	
Db	560	AAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	619	
Qy	727	actaatctcttc-aagagtatgtgataaagaacgctatactacagcaagagcctcttaa	785	
Db	620	TCAATTTTTTTTTTWWAATAAAARWTCGAARGGKGGGGKTTTKGGATGRRRCAATA	679	
Qy	786	atatttggactaaaatctaccocctagtagtattactattagaaaaactatatttggttata	845	
Db	680	AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	739	
Qy	846	tttatactttaagtcctccttggttagagggttattataataatgatatataatttagt	905	
Db	740	KRAAARDWAATKRAATRTKRTKRDWTKDNGNWTGAAAAATTTTTRDRRTTWTTRDKKRA	799	
Qy	906	actgaaattggttgaggagagaacaaa	934	
Db	800	KTDWNCATTTTDRRAATTKRAAAAAAR	828	
RESULT 6				
LOCUS CNS0102F/c	928 bp DNA linear GSS 26-JUL-1999			
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN03B09 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				

```
ACCESSION AL098433.1 GI:5610044
VERSION AL098433.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 928)
AUTHORS GSS.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Osoegawa and
          Aaron Mamoser in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCI-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
          pl and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source           1..928
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BAC27A24"
                     /note="end : T7"

BASE COUNT  169 a   128 c   100 g   241 t   290 others
ORIGIN

Query Match      7.3%; Score 68.8; DB 12; Length 928;
Best Local Similarity 24.5%; Pred. No. 0.046;
Matches 126; Conservative 156; Mismatches 232; Indels 0; Gaps 0;

Qy 6 ttaagtagtattatgtcggtatataatgagcctttaaattatgtgagattcagtaga 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 924 TAAAGKAGWAGWRRWAATDRWATAWTAKWRWAGCTRWATARKATKRGAGAKRWAA 865

Qy 66 atctattataaacaacgcttactgatttttgagttcataattgcttataatcccaag 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 864 RWATAKAKWGTWCAKWTADKDKTGGTGTGWTGTRGTRGKRRRAGGDTGRGDRTR 805

Qy 126 tagaggtgatttaagaacaattcttaacagaataattcagttgtagataatagaataaaaaat 185
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 804 TRRAGKAGRWAKARAKTAWAADNGARWAWAAAWADATGAGATAWAWATAWTATAAKATA 745

Qy 186 ctgcttaataagaagaaataattggttttagcatcaagtttgacaagcggtgaaatttc 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 744 TTAWAWRRDRARATGTGKWRGTGTTAKAGATKGGWDDTRADAKKAAGGTTATAKTATA 685

Qy 246 taegggagaatattttttagaattggtgctgatgatttcatattcacaagtagattga 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 684 TGTARARAKAKWKAATAKAKAGATATRAAWATGADAWGATWTGKKTAGAKTTGTAGW 625

Qy 306 taagcaaatctgtttatgaggagaaattcattggtattcttcacgaactctaatagaatt 365
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 624 TGRKWTDKRRKTDKRDATGWRRAATAWAGAADDAWDRGWAGTCAGRRRRDGRWRARA 565

Qy 366 gatagacccaaaagaaatttagTataataaacaacgagaaagtaaaaaataatactaaac 425
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 564 DTRKAGADRRAAAAAAGTGTGADAADKRAARGKDKGDKRKKTKGKRAATGGWRGAG 505

Qy 426 taatgatatacgggaagtgttattgaatagatctctactgtgcccccacacgctggtcgct 485
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 504 RAKDGRKRGWKKKRAKDKTKTKWDTTRTDKAKAKATGAKDGAKGKRWAGDKRKKGGG 445

Qy 486 aaaaaagaagtttctcgataagtttaattgggatat 519
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 444 AGATATRTGKGGWDDWATTTGAWTTATGRTATAT 411
```


BASE COUNT 302 a 83 c 32 g 379 t 305 others
ORIGIN

Query Match 7.3% Score 68.2; DB 12; Length 1101;
Best Local Similarity 31.1%; Pred. No. 0.056;
Matches 276; Conservative 160; Mismatches 441; Indels 10; Gaps 2;

```
Qy 40 ttaaatattgagagattcagtagaattctataataatcaaacgcttactgatttgag 99
Db 1084 TTTTTHAMMAAAWYWAACACVACAHHMAWAAWAAAHYCTCTTATTATTWTG 1025
Qy 100 ttcataattgctattgataaccagtagagtgatttaagcaattcttaacgaatat 159
Db 1024 KSANSAHKTGCKKTTTTTKKAAWTKTATAAAVAANGRMAAARAANGKAATM 965
Qy 160 tcagttgtagataatagaataaaattctgcttaataagaataatattggttagcatca 219
Db 964 AKKTKTATTAATYTTTWTYKBDTTTTTTTTTAAASMYWRGKKTAAADKTGAATTTA 905
Qy 220 agtttgacaaacggtgaaattcttaagggagaataatttttagaattgagctgat 279
Db 904 RKBKACAGAAAAAAGCGSAMMBTACKGKAGCKCKCMAADTWKTTWATTKAAAK 845
Qy 280 gatatttcataccaagtagatttgataagcaaaattcgtttttagggagaaattcattg 339
Db 844 ACTATTAKKSMGSKAGAAKRGCKDAATAAATAAKSKGKGYCKMCAAWKDGAAAMK 785
Qy 340 gattttcagcaactctaataagaat-----tgatagacaaacgaaggaatttagtat 391
Db 784 KKSACAAKAAAKRAWAAGKKKAKAAGGKAGKGRAGKAAHAAADAAAAAAGGWAAGKAKK 725
Qy 392 ataacaacagagaaagtaataataatacttaactaactatgatatcaggaagtgattga 451
Db 724 GGGAAACGKKKAAAGDAAKGAHAAHAAHAAHAAHAAHAAHAAHAAHAAHAAHAAHAAH 665
Qy 452 atagattatacttgcacccaacggtggtcgtaaaaaaagaagtttctgataagttaa 511
Db 664 KKMAGRAKAKACAAKAAAGTKAGACACAKAKAAAAAAGDCKKMGKDKAKAAARAKMKAA 605
Qy 512 tgggatatagagatttgtagtactgtgaagattatgatttgcataaagagagctctgg 571
Db 604 AKKAKAKKAKKKKAKKAAKDWCKRAAKKWAHAAKDKKAGAGAADKKGKAAAAAGAAK 545
Qy 572 ctgatttcaaaatcgcttactcctaagaagctattttacagtatagattaaacagaatg 631
Db 544 AKAATATAMSGATTGKWCAGKATAAGAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 485
Qy 632 gaatatcacaacccaataaagtttaagcaatatatttactcagctattttcaagaatttt 691
Db 484 AAKATGAATATAATTTAAATGTTTTTGTATTATTATTTTTTTTATTATATATTATTTT 425
Qy 592 ataaagaaaaatcttatattgataacacaaaaatctactaaatctacttcaagagatgta 751
Db 424 TTAATATAAAAKTATTTTGAAG--ATTAAAGTATTTTTTTTATTATWATWATATATTA 367
Qy 752 taaagaacgctatactacagagagctctcctaataatttttgagctaaatctaccctta 811
Db 366 KTWATATAAAKAAWAAWAAWWTAAWAAATWAAATATTATTTTAAAGTTTTTTTTT 307
Qy 812 gtattactattagaaaaactatatattgtttatatatttatactttaaagctcccttggtta 871
Db 306 TTWTTTAKATAATAATTTTTTTTAAATTTTTTTTTTTTTTTTTTTTTTTTWTATWATATA 247
Qy 872 ggaggttatttaaatagatatttaatttttagtactgaaattgttt 918
Db 246 WAAATTAATWATWATKWATTKWATTDATTATATTATTTGTGTTTWTGTTTTTTTTT 200
```

RESULT 10
CNS01400
LOCUS CNS01400 1001 bp DNA linear GSS 26-JUL-1999

DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN11B06 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL103554
VERSION AL103554.1 GI:5615165
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 1001)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segrefgenoscope.cns.fr
- web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES
source Location/Qualifiers
1..1001
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN11B06"
/note="end : SP6"

BASE COUNT 263 a 70 c 145 g 236 t 287 others
ORIGIN

Query Match 7.2% Score 67.2; DB 12; Length 1001;
Best Local Similarity 24.8%; Pred. No. 0.082;
Matches 105; Conservative 131; Mismatches 187; Indels 0; Gaps 0;

```
Qy 2 tgaattatagtcattatgctcggtatataatgagcctttaaattatgtgagagattcag 61
Db 563 TNNNTTANAWTATNTNTTWTNTTNTYASTAAGAAVTMBMBANGANRCNNHATBTNKT 622
Qy 62 tagaatctataataatcaaacgcttactgatttgaattcatttaatttcattgataatc 121
Db 623 DRRGTGKWKDTRABAWARRRKAATKRWTTKTKTDAGAKATADAAWAAVAKWAKGTNGW 682
Qy 122 caagtagagggtgatttaagcaattcttaacagaattctcagttgtagataagaataa 181
Db 603 WAWWAWYAKADTAKADTDWTADTGTWGTAKAWADWTTKWANWKRAGGKRKAKDKR 742
Qy 182 aaatctgtcttaagaagaaataattggttagcatcattgagtttgacaaagcggcgaaa 241
Db 743 KDKWDKWKARTGKGGGTWDKDKNRKAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
Qy 242 ttttaaggggagaataattttttagaattggatgctgtagatatttcattcattcaagtag 301
Db 803 TATMNKAWHWTABTAKATAAAWADTAKAAWGDGWAANAATDKDWTWKKADADTTATAT 862
Qy 302 ttgataagcaaatctgtttttagggagaaattctcattgatttctcagcaactcaatag 361
Db 863 AKKWWABABAKAHDPTAAWADKWAARNAADATTATAWADNWKAKWAGKAKABAD 922
Qy 362 aattgatagacaaaggaatttagtatataaaacacgagagaaagtaataaaaaataact 421
Db 923 HDKGGADAKADAKAWAKAAAWTAAAAADWAKAGAKAWATADTAKTADADTAKAKADT 982
Qy 422 taa 424
Db 983 DTA 985
```


Qy	759	acgctatactcagcaagagcgctctctaaatattttgagctaaaatctccccctagtattac	818
Db	795	A---TAAATATATTATAATTATAATAAATAAAATATATATTATTAATAATAA	851
Qy	819	tattagaaaaactatatatttggttatatttatactttaagctcccccctggtaggaggtt	878
Db	852	AAATATTAAATTTNTTTTATTAAATAAAANTNTNATAATATATAATAAT	911
Qy	879	attaataaatgatattaatatatttt	902
Db	912	ATAATTTAAAATATATTAAATATTTTT	935

RESULT 15
CNS020K7

RESOLUTION	LOCUS	DEFINITION	1092 bp	DNA	linear	GSS 12-MAY-2000
CNS020K7	CNS020K7	Tetraodon nigroviridis genome survey sequence T7 end of clone 222L11 of library G from Tetraodon nigroviridis, genomic survey sequence.				
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						

REFERENCE
1 (Pages 1 to 102)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

[illegible]

```
QY 263 ttgaatggatgctgatgatttcattccaaagtagatttgataagcaaatcgttta 322
Db 746 AWWNAWTTTWTTTTWWWWWWWWAAAAAANWNAWTTTAAWTTTAAWTTTAAWTTTAA 805
QY 323 tggaggaaaatttcattggatttctcagcaactctaatagaaattgtagaccacaaaggaa 382
Db 806 TTTTAAANAAAAANWTTTAAWTTAATAAAAAANWNAWTTTAAWTTTAAWTTTAAWNA 865
QY 383 atttagtatataaacaacgagaagtaataaaatatataacttaactaatgatatacggaaga 442
Db 866 TTWNAATWTAATAAAAAAATAAAAAAANWTTTWTTTTWTWNAWTTTAAATWTTT 925
QY 443 tgttattgaatagatctatacttgccaccaccaacgctggtcgtaaaaaagaaagtttcg 502
Db 926 TTTTWTTTTWTWTTTAAWNAAAAAAATAAAAAAATAAAAAAANWTTTAAATTTATTT 985
QY 503 ataagttaatggagatatagagatttagtacctgttgagatttatgattttgcaataaga 561
Db 986 TTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 1044
```

Search completed: August 2, 2002, 12:56:53
Job time: 8219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 12:22:04 ; Search time 210.05 Seconds
(without alignments)
7675.233 Million cell updates/sec

Title: US-09-900-038A-2

Perfect score: 939

Sequence: 1 atgaattatagatcatcatt.....gaggagagaacaaagtgc 939

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	70.2	7.5	15548	24	ABL34155
C 2	68.6	7.3	5371	24	ABL34294
C 3	68.2	7.3	14024	20	AA06749
C 4	68.2	7.3	14024	21	AAZ56386
C 5	67	7.1	4915	20	AA013062
C 6	64.8	6.9	906	21	AA03723
C 7	64.8	6.9	11474	21	AA03720
C 8	64	6.8	7597	24	ABL33013
C 9	63.8	6.8	1170	21	AA03728
C 10					Human immune syste
C 11					Human immune syste
C 12					E. coli O157 anti
C 13					Escherichia coli f
C 14					Enterococcus faeca
C 15					Campylobacter jeju
C 16					LPS core biosynthe
C 17					Human immune syste
C 18					Campylobacter jeju

C 10	63.8	6.8	11474	21	AA03720	LPS core biosynthe
C 11	63.4	6.8	5154	20	AA013138	Enterococcus faeca
C 12	63	6.7	5689	22	AA045384	Chemically pretrea
C 13	63	6.7	5689	22	AA046426	Tumour suppressor
C 14	60.4	6.4	912	21	AA03727	Campylobacter jeju
C 15	60.4	6.4	7814	22	AA046530	Tumour suppressor
C 16	60.2	6.4	12237	24	ABL34358	Human immune syste
C 17	59.8	6.4	700	22	AA093026	Human inflammatory
C 18	59.6	6.3	6286	22	AA046591	Tumour suppressor
C 19	59	6.3	8370	22	AA046713	Tumour suppressor
C 20	58.8	6.3	2118	23	AA051313	Enterococcus faeca
C 21	58.8	6.3	2148	23	AA052777	Enterococcus faeca
C 22	58.8	6.3	4054	20	AA013188	Enterococcus faeca
C 23	58.6	6.2	6474	19	AA052288	Streptococcus pneu
C 24	58	6.2	6109	24	ABL32326	Human immune syste
C 25	58	6.2	6109	24	AA061077	Human gene regulat
C 26	57.8	6.2	3991	22	AA016633	Human novel protei
C 27	57.8	6.2	6992	21	AA060930	Nucleotide sequenc
C 28	57.8	6.2	9539	22	AA045346	Chemically pretrea
C 29	57	6.1	18183	22	AA046280	Tumour suppressor
C 30	56.6	6.0	15950	21	AA045260	DNA encoding enzym
C 31	56.6	6.0	15950	21	AA030357	Nucleotide sequenc
C 32	56.6	6.0	19124	18	AA072882	Plasmodium var-7 g
C 33	56.6	6.0	19124	21	AA098287	Plasmodium var-7 p
C 34	56.4	6.0	1671	13	AA024134	50 kD subunit of S
C 35	56.4	6.0	37973	24	ABL34197	Human immune syste
C 36	56	6.0	18218	24	ABL33949	Human immune syste
C 37	55.6	5.9	5883	24	ABL33733	Human immune syste
C 38	55.6	5.9	19124	18	AA072882	Plasmodium var-7 g
C 39	55.6	5.9	19124	21	AA098287	Plasmodium var-7 p
C 40	55.4	5.9	6154	24	AA061112	Human gene regulat
C 41	55.2	5.9	700	22	AA093026	Human inflammatory
C 42	55	5.9	20420	22	AA073165	Human immune/haema
C 43	54.8	5.8	5930	24	ABL32517	Human immune syste
C 44	54.8	5.8	6419	24	ABL32267	Human immune syste
C 45	54.8	5.8	11155	24	ABL32605	Human immune syste

ALIGNMENTS

RESULT 1
ABL34155/c
ID ABL34155 standard; DNA; 15548 BP.
XX
AC ABL34155;
XX
DT 26-MAR-2002 (first entry)
XX
Human immune system associated gene SEQ ID NO: 2128.
XX
Human; immune system disease; cytosine methylation; antiasthmatic;
XX
antiarteriosclerotic; antianaemic; cytostatic; nontropic;
XX
neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX
antihematuric; antithrombotic; antidiabetic; antipsoriatic;
XX
antihematuric; cancer; eye disease; arteriosclerosis; anaemia;
XX
acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX
neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX
gene; ds.
XX
Homo sapiens.
OS
XX
WO200200928-A2.
PN
XX
03-JAN-2002.
PD
XX
02-JUL-2001; 2001WO-EP07537.
PF
XX
30-JUN-2000; 2000DE-1032529.
PR
XX
01-SEP-2000; 2000DE-1043826.
PR
XX
(EPIC-) EPIGENOMICS AG.
PA
XX

PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 2128; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 15548 BP; 4209 A; 247 C; 2903 G; 8189 T; 0 other;
 XX
 Query Match 7.5%; Score 70.2; DB 24; Length 15548;
 Best Local Similarity 44.3%; Pred. No. 0.00037;
 Matches 332; Conservative 0; Mismatches 413; Indels 4; Gaps 1;
 QY 173 atagaataaaatctgcttaagagaataatttttagaatggatgctgatatttcataac 232
 DB 12140 ATATATACATATAAATAACATATAAATAACATATAAATAACATATAAATAA 12081
 QY 233 cgggtgaaattcttaagggaggaatatttttagaatggatgctgatatttcataac 292
 DB 12080 CATATAAATAAATACATATAAATAAATACATATAAATAAATATAAATAA 12021
 QY 293 caagtagattgataagcaaatctgttttagggaggaataattcattggttctcagaa 352
 DB 12020 AATAAATAATATATACATATAAATAAATATAATATATATACATATAAATA 11961
 QY 353 ccttaataagattgataagcaaaagaaatttttagtatataaacaagagagaataca 412
 DB 11960 ATAACATATAAATAACATATAAATAAATATAAATAAATAAATAAATAAATA 11901
 QY 413 aaataacttaactgatatacaggaagatgttttagaatgatctctacttgcaccacc 472
 DB 11900 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 11841
 QY 473 caacgtggtgcataaaagaagtttcgcataagtttaagggatagagatttagtac 532
 DB 11840 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 11781
 QY 533 ctgttgaagattgatgttttgcaataagaggagctgctgctgatttcataaatcggttac 592
 DB 11780 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 11721
 QY 593 tcaataaagttacttttacagtagatataacgagaatgggaatatacacaacccaataagt 652
 DB 11720 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 11661
 QY 653 ttaagcaatattactactcagctattttacaagattttttataaagaataatcttatattg 712
 DB 11660 ATAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 11601
 QY 713 atatacaaaaattactaattcttcaagagtagtgataagaacgcctactactcagc 772
 DB 11600 TAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 11541
 QY 773 agagctctcctaaatatttttagagctaaaatctacccttagtattactattagaaactat 832
 DB 11540 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 11481
 QY 833 atatttg----tttattattactttaaagtcctcctggttaggaggttattataaat 888
 DB 11480 ATATATTCACATTACATTCTTTTACAAACATTCATTTAATTTATATAATTAATA 11421

QY 889 gatattaataatttttagtactgaattgtt 917
 DB 11420 TACTTAATTATTATAACACGAAATATTT 11392
 RESULT 2
 ABL34294/C
 ID ABL34294 standard; DNA; 5371 BP.
 XX ABL34294;
 XX 26-MAR-2002 (first entry)
 DT Human immune system associated gene SEQ ID NO: 2267.
 DE Human; immune system disease; cytosine methylation; antiasthmatic;
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antiarteriosclerotic; antiarteriosclerotic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS WO200200928-A2.
 XX 03-JAN-2002.
 PD 02-JUL-2001; 2001WO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 2267; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 5371 BP; 1322 A; 87 C; 1312 G; 2650 T; 0 other;
 XX
 Query Match 7.3%; Score 68.6; DB 24; Length 5371;
 Best Local Similarity 47.3%; Pred. No. 0.0007;
 Matches 239; Conservative 0; Mismatches 264; Indels 2; Gaps 1;
 QY 227 acaagcgggtgaaatttctaaggaggagaataatttttagaatggatgctgatgatatt 286
 DB 1804 AAAACTCTATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1745
 QY 287 catatccaagtagatttgataagcaaatctggttttagggagaaatctcattggtttct 346
 DB 1744 TAATCAATATCTTTTTTAAAAAATAAATAAATAAATAAATAAATAAATA 1685
 QY 347 cagcaactctaataagaattgtagaccacaaaggaatttagtatataaacaacgagaaa 406

Matches	198;	Conservative	0;	Mismatches	173;	Indels	6;	Gaps	2;
---------	------	--------------	----	------------	------	--------	----	------	----

Qy	13	atcattatgtcgttatataatagagcctttaaatatgtgagagattcagtagaattctata	72
Db	100	ataattatgccgtttacaatggggc---caaaactataatctcatcagtagaatcaatt	156
Qy	73	ttaaatcaaacgcttactagatttttgagttccataattgtcattgataatcccaagtagaggt	132
Db	157	atacatcaactttatcaagattttg---ttttgtatatcattgacgattgtagaccgat	213
Qy	133	gattttaaagcaaatctttaacagaatattcagttgtagataatagaataaaaaattctgctt	192
Db	214	gatacttttcttaatacaacagtcgatacaaaaacaatcagaaaataagaataattgcgt	273
Qy	193	aatgaagaaaaatatgtgttagcatcaagtgttgacaaagcgtgaaaaattcttaaggga	252
Db	274	aacaagacaaatttagtgtgttcgaaagtcgaattatggaatagaataatggccacgggg	333
Qy	253	gaatatattttgaatggatgctgatgatatttccatcccaagtagattgataagcaa	312
Db	334	aaatatattttttgtgatgcggatgatttggcagcagaaaaaatttagcgtcaaa	393
Qy	313	attcgttttatggaggaaaaattccattgattttctcagcaactctaaatagaattgatagac	372
Db	394	atcgaagtgttaataatgaatgtgtagatggtgttcttaattattatgttatagat	453
Qy	373	caaaaaggaaaatttagt	389
Db	454	aacaatagaaaatatgt	470

RESULT	4
AAZ56386	
ID	AAZ56386 standard; DNA; 14024 BP.
XX	
AC	AAZ56386;
XX	
DT	17-MAR-2000 (first entry)
XX	
DE	Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.
XX	
KW	Flagellin; fliC; antigen; detection; ds.
XX	
OS	Escherichia coli.
XX	
PN	WO961458-A1.
XX	
PD	02-DEC-1999.
XX	
PF	21-MAY-1999; 99WO-AU000385.
XX	
PR	21-MAY-1998; 98AU-0003634.
XX	
XX	(UNSY) UNIV SYDNEY.
PA	
XX	
PI	Reeves PR, Wang L;
XX	
XX	WPI; 2000-072598/06.
DR	
XX	

Novel nucleic acid molecule useful for the detection of flagellated bacterial strains in food, faeces, etc.	-
Claim 3; Page 229-233; 245pp; English.	

AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or part of an Escherichia coli flagellin protein except a protein expressed by E. coli H1, H7, H12 or H48 type strains. The present invention also describes a method of detecting the presence of E. coli of a particular H serotype in a sample, comprising specifically hybridising a nucleic acid, preferably at least a pair, derived from a flagellating gene, specific for a particular flagellin gene associated with the H serotype, to any E. coli in the sample which contain the gene, and detecting any hybridised molecules, identifying the presence of that	
--	--

CC	serotype in the sample. (I) are useful for: (1) detecting the presence						
CC	of E. coli of H serotype in a sample by hybridising at least one or a						
CC	pair of (I) to any E. coli in the sample and detecting the hybridised						
CC	nucleic acid molecules; and (2) for detecting the presence of both O						
CC	and H-serotypes of E. coli by hybridising at least one or a pair of (I)						
CC	to any E. coli present in the sample and detecting the hybridised						
CC	nucleic acid molecules. (1) is particularly useful for detecting the						
CC	combination of O and H antigen. Hybridised (I) when using at least one						
CC	(I) is detected by southern blot analysis and, when using a pair of (I),						
CC	is detected by polymerase chain reaction (PCR). AA256399 to AA256420						
CC	represent primers used in the exemplification of the present invention.						
XX							
SO	Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 other;						
	Query Match	7.3%	Score 68.2:	DB 21;	Length 14024;		
	Best Local Similarity	52.5%	Pred. No. 0.00084;				
	Matches 198;	Conservative 0;	Mismatches 173;	Indels 6;	Gaps 2;		
Qy	13	atcattatgtcggtataataagacgcctttaaatatgtgagagatccagtagaatctata	72				
Db	100	ataattatgccggtttacaattgggc---caaaactataatcatcagtagaatcaatt	156				
Qy	73	ttaaatcaaacgcgttactgatttttgatttcataaattgtcatttgataatccaagttagagg	132				
Db	157	atacatcaatcttatcaagattttg---tttgttatcatcattgcagattgtaccgcgat	213				
Qy	133	gattitaaagcaattcttaacagaatatctcagttgtagataataagaataaaaaatcttgctt	192				
Db	214	gatacatcttcttaatacaacagtcgatacaaaaacatcagaaaaaagaatattgcgt	273				
Qy	193	aatgaagaaaaatatggtttagcatcaagtttgaacaaagcgttgaaatcttctaagggga	252				
Db	274	acaagcaaaatttaggtgtgcagaagtcogaatttatggaatagaaatggccacgggg	333				
Qy	253	gaatatatttttagaattggtgctgatgatatatttcataatccaagtagatttgataagcaa	312				
Db	334	aaatatattctttttgtgtagcggtgatttggcgacgagaaaaaattagagcggtcaa	393				
Qy	313	attcggtttatggaggaaaaattcattgatttcttcagcaaacactctaatagaattgataagac	372				
Db	394	atcgaaagtctaaataatgaatgtgtagatggtgtatgttcttaattattatgttatagat	453				
Qy	373	caaaaaggaaaatttagt 389					
Db	454	acaatatgaatatattgt 470					
RESULT	5						
ID	AAAX13062/C						
XX	AAAX13062 standard; DNA; 4915 BP.						
AC	AAAX13062;						
XX							
DT	19-MAR-1999 (first entry)						
XX							
DE	Enterococcus faecalis genome contig SEQ ID NO:125.						
XX							
KW	Enterococcus faecalis; contig; detection; Enterococcal infection;						
KW	vaccine; attenuation; computer readable medium; ds.						
XX							
OS	Enterococcus faecalis.						
XX							
PN	WO9805055-A2.						
XX							
PD	12-NOV-1998.						
XX							
PF	04-MAY-1998; 98WO-US08985.						
XX							
PR	14-NOV-1997; 97US-0066009.						
PR	06-MAY-1997; 97US-0044031.						
PR	16-MAY-1997; 97US-0046655.						
XX							

Db 5722 GTAGCAAACTCTCCTTATATAAATCTTTTATAGATCCTGATGATTATTGGAACTAAATGCT 5663
 QY 301 ttgataagcaaatctgtttttatggagagaaaa 332
 Db 5662 TGTCAGAGTGTATAAAATTTTATAGATGAACA 5631

RESULT 8
 ABL33013/C
 ID ABL33013 standard; DNA; 7597 BP.
 XX
 AC ABL33013;
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 986.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosstatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 XX WPI; 2002-130909/17.
 XX
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 986; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 other;

Query Match 6.8%; Score 64; DB 24; Length 7597;
 Best Local Similarity 45.2%; Pred. No. 0.0048;
 Matches 313; Conservative 0; Mismatches 375; Indels 4; Gaps 2;

QY 167 tagataatagaataaaaaattctgttaataagaagaaattgttttagcatcaagtttga 226
 Db 866 TAAATAATAAAATAACATAAAATTAATACATAAAATTAATAATAATAATAATAATAA 807
 QY 227 acaagcggtgaaattcttaaggaggaatatttttagaatggatgtagatattt 286
 Db 806 TAAATAATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAT 747

QY 287 catatccaagtagatttgataagcaaat-tcgtttttttgaggagaaattcattgtgatttc 345
 Db 746 AACGCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAC 587
 QY 346 tcagcaactctaatagaaattgtagagacccaaagaaatttagtatatacaaacagagaa 405
 Db 686 AAATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 627
 QY 406 agtaataaaatatataacttaactaatagatacggaagatgttatgaatagatctatactt 465
 Db 626 ACGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 567
 QY 466 gccccaccaacgtgtgcgttaaaaaagaaagtcttcgataagtttaagtggatagagat 525
 Db 566 AAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 507
 QY 526 ttgtaacctgttgagattatgtttgcaataagagagagctcgtgctgatttcaaaatc 585
 Db 506 AAATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 447
 QY 586 ggcctactcaataaagctacttttacagtatagattaaacgagaatggaatcacaaacc 645
 Db 446 AAATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 387
 QY 646 aataagtttaagcaatatatttactcagctattttacaagatttttataagaagaaatct 705
 Db 386 AAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 327
 QY 706 tatattgatatacaaaaaattacttaattacttcaagagtgatgtgataaagaacgctat 765
 Db 326 AAATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 270
 QY 766 actcagcaagagctctctaaatatatttgagctaaatactaccctagttactattaga 825
 Db 269 ATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 210
 QY 826 aaactatataattgttttatatttatactttta 857
 Db 209 CGAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 178

RESULT 9
 AAA53728
 ID AAA53728 standard; DNA; 1170 BP.
 XX
 AC AAA53728;
 XX
 DT 22-DEC-2000 (first entry)
 XX
 DE Campylobacter jejuni OH4384 galactosyltransferase.
 XX
 KW Biosynthetic locus; biosynthesis; lipid A biosynthesis;
 KW acetyltransferase; glycosyltransferase; Beta-1,4-GalNAC transferase;
 KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
 KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
 KW immunity; immunogen; ganglioside; ds.
 XX
 OS Campylobacter jejuni OH4384.
 XX
 FH Key Location/Qualifiers
 CDS 1..1170
 FT /*tag- a
 FT /product= CstII sialyltransferase
 FT
 XX WO2000046379-Al.
 PN
 XX 10-AUG-2000.
 PD
 XX
 PF 01-FEB-2000; 2000WO-CA00086.
 XX
 PR 01-FEB-1999; 99US-0118213.
 PR 31-JAN-2000; 2000US-0495406.

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 10:47:14 ; Search time 49.3 Seconds
(without alignments)
4678.496 Million cell updates/sec

Title: US-09-900-038A-2

Perfect score: 939

Sequence: 1 atgaattatagatcattat.....gaggagagaaacaaagtgc 939

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	56.6	6.0	19124	2	US-08-487-826B-13
C 2	55.6	5.9	19124	2	US-08-487-826B-13
C 3	51	5.4	854	4	US-08-998-416-534
C 4	50	5.3	860	4	US-08-998-416-287
C 5	50	5.3	6124	4	US-08-213-419B-3
C 6	49.4	5.3	7218	1	US-08-232-463-14
C 7	48.8	5.2	1882	1	US-08-257-073-12
C 8	48.8	5.2	1884	1	US-08-257-073-8
C 9	48.6	5.2	615	4	US-08-998-416-186
C 10	47.6	5.1	844	3	US-08-961-083-195
C 11	47.4	5.0	837	4	US-08-998-416-288
C 12	46	4.9	3701	4	US-08-845-258-10
C 13	46	4.9	3701	4	US-08-990-571-10
C 14	46	4.9	3701	4	US-08-723-142A-10
C 15	45.6	4.9	636	4	US-08-998-416-1137
C 16	45.6	4.9	665	2	US-08-883-795A-36
C 17	45.6	4.9	6152	4	US-08-973-462-1
C 18	45.6	4.9	8920	2	US-08-446-855A-1
C 19	45.6	4.9	8920	4	US-09-150-741-1
C 20	45.4	4.8	658	4	US-08-998-416-595
C 21	44.4	4.7	837	4	US-08-998-416-288
C 22	44.2	4.7	732	4	US-08-998-416-1036
C 23	44	4.7	724	4	US-08-998-416-683
C 24	43.8	4.7	688	4	US-08-998-416-972
C 25	43.8	4.7	1684	4	US-08-936-165A-259
C 26	43.4	4.6	782	4	US-08-998-416-224
C 27	43	4.6	767	4	US-08-998-416-472

Query Match 6.0%; Score 56.6; DB 2; Length 19124;

ALIGNMENTS

RESULT 1

US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; APPLICANT: Chitnis, Chetan

; APPLICANT: Miller, Louis H.

; APPLICANT: Peterson, David S.

; APPLICANT: Su, Xin-zhaun

; APPLICANT: Wellem, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,826B

; FILING DATE: 10-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Israelsen, Ned

; REGISTRATION NUMBER: 29,655

; REFERENCE/DOCKET NUMBER: NIH121.001CPI

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 19124 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: Single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; US-08-487-826B-13

Sequence 191, App
Sequence 191, App
Sequence 32, Appl
Sequence 32, Appl
Sequence 4, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 1092, Ap
Sequence 224, App
Sequence 541, App
Sequence 534, App
Sequence 287, App
Sequence 487, App
Sequence 1, Appl
Sequence 1, Appl

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meligs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 287:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 860 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1240UP
US-08-998-416-287

Query Match 5.3%; Score 50; DB 4; Length 860;

Best Local Similarity 45.5%; Pred. No. 0.012;
Matches 252; Conservative 0; Mismatches 300; Indels 2; Gaps 2;

```
QY 74 taaatcaaacgcttactgattgttgatcgaattgtcgaatgaatccaagtagagtg 133
DB 616 TAAAGGAAATCGGAATATTTGGCACTTAAATTTTATTTATTTAAATGTTATCTA 557
QY 134 atttaaacgaattcttaacagataattcagttgttagataatagaataaaacttgctta 193
DB 556 TTTAAACATAAAACATTTTAAATGTTATAAAATAAATAAGAAATTACTTATAGATATT 497
QY 194 atgaagaaaatttggttttagcatcaagtttgaaacaaagcggtgaataatcttaaggag 253
DB 496 ATTAATAGTATTTAAATTTAAATTTTAAATTAATATACCATTTTATTAATAAATAGAT 437
QY 254 aaatatttttagaaggatgctgatgatattcatatccaagtagatttgataagcaaa 313
DB 436 TATTAGTTTATATTAATTAAGTGATATATAATTTAAATTTATATAAATTTAATTTAC 377
QY 314 ttctgtt-tatggagaaaattcattggtttctcagcaactctaatagaattgatagac 372
DB 376 TTCATTGATATATAAATTTAAATGTACCTTTTCATAATATTTATTTTATTAGTCTAG 317
QY 373 caaaaagaaaatttagtatataaaacacagagaagaataataataacttaactaatgat 432
DB 316 TAAATTTCTATTATTAGTCTACCTTTAATGGATATTACTACTATAAATTTTAC 257
QY 433 atacggaagatgtattgaatagatctataacttgcaccccaacggtggtgcgtaaaaag 492
DB 256 CTA-ATAATATATTATAGATATACTTAATCTAATAATTTTATTATCTAAAGTATATAA 198
QY 493 aaagtttcogataagttaatgggatagagatttagtaccttggttgaagattatgatttt 552
DB 197 TTAATTAATCTTTTATTATTATTAAATTTATTATTAAATTTAGTAAATTTATATTATT 138
QY 553 gcaataagaggagctcggtgatattcaaaatcggtctactcaataaagtaacttttaag 612
DB 137 ATTTATTAACTAATATTTTGTATAATAATATATCATTTAAATGTTAAATTTTAAATA 78
QY 613 tatagattaaacga 626
DB 77 ATTATCTTTAATGA 64
```

RESULT 5

US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE REFERENCE: JII-002CNCp
; CURRENT APPLICATION NUMBER: US/08/213,419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match 5.3%; Score 50; DB 4; Length 6124;

Best Local Similarity 44.3%; Pred. No. 0.016;
Matches 203; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

```
QY 298 agatttgataagcaaatcgtttttatggaggaaaattcattggatttctcagcaactcta 357
DB 2169 aaattcaataaatttaataataataataattttatattataatttataatatataata 2228
QY 358 atagaattgatagacaaagaaaatttagtatataaaacacagaaagaaataataataa 417
DB 2229 ataaatttttttaatttaaaatttaaatttagattgtccaaaaaaataaaaaataaat 2288
QY 418 tacttaactaatgatatacgggaagattgttattgaatagatctatacttgcacccaacg 477
DB 2289 atatatatatataaaatacatatatattataacataaagaaaataataataatacaac 2348
QY 478 tgggtgcgtaaaaagaaagtttctcgataagtttaattgggatatagagatttagacctgt 537
DB 2349 atattcaaaaaaattaaagttcttaaaattattataataaacttaattcatatcaaaat 2408
QY 538 gaagatttatgatttgcataaagaggctctggtgatttcaaaatcggttactcaat 597
DB 2409 gaagtcatatattctctgttttttcataatcgtaagaatgaaaaaaataaaaaaagga 2468
QY 598 aaagtacttttaccgtatagattaaacgagaatggaatatcacaaaccaataagtttaag 657
DB 2469 aaagaaaaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 2528
QY 658 caatatatttactcgctgattttacaagaatttttataaagaaaaattcttatattgataac 717
DB 2529 ataatattataataataataataatttttttcgcatacacacaaacatttgcattat 2588
QY 718 acaaaaattactaatcttcaagagtatgtgataa 755
DB 2589 tttttttttagtgcataatttcaaaaaatgttataa 2626
```

RESULT 6

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/232,463
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 5.3%; Score 49.4; DB 1; Length 7218;
Best Local Similarity 9.1%; Pred. No. 0.022;
Matches 38; Conservative 200; Mismatches 181; Indels 0; Gaps 0;
Qy 123 aagtagaggtgatttaagcaattcttaacagaaatattcagttgttagataataagaataaa 182
Db RRR 1402
Qy 183 aattctgttaagaagaataattggttagcatcaagttgaacaaagcgtgaaat 242
Db RRR 1342
Qy 243 ttctaaggagaaatatttttagaagtgatgctgagataatttcataccaagtagatt 302
Db RRR 1282
Qy 303 tgaagaacaattcttttagaggaataattcattgattctcagacaactcaataga 362
Db RRR 1222
Qy 363 attgatgaccaaagaatttagtatataaacaacgagaaagtaataataatactt 422
Db RRR 1162
Qy 423 aactaatgataacggaagatgttatgaatagatctatacttcccaaccccaacggtg 482
Db RRR 1102

Qy 483 cgtaaaaaaagaagtttttcgataaagtttaattggtgatagagatttagtacctgttgag 541
Db 1042 CCAAGCTCGGAATTAATTCGTGAGCGTATGCGCAACGAAGAAAAATAGTTATAGTAG 984
RESULT 7
US-08-257-073-12
Sequence 12, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1882 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-12
Query Match 5.2%; Score 48.8; DB 1; Length 1882;
Best Local Similarity 47.4%; Pred. No. 0.024;
Matches 146; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
Qy 106 attctatctgataccaagtagagggtgatttaagcaattcttaacagaataattcagtt 165
Db 466 ATAATATTGAGAAATCAATFACTACTTTTATACACCGGTAGCTACCGGAAATCAATAT 525
Qy 166 gtagataataagaataaaaaattctgcttaataagaagaaatttggttttagcatcaagtttg 225
Db 526 TTAAGATGAGGTTTTCCTCCCAACAGAACCTCTTATGTCACCAATGACATTA 585
Qy 226 aacaaagcgtgaaattcttaaggagagaatatatttttagaagtgatgagtagattt 285
Db 586 GATGAAATGAGACATTTTATATAAGATAAATAATATGTAATAAATTTAGATGAATTGACT 645

Qy	286	tcatatccaagttagattgataagcaaaattcgtttctatggaggaaaaattcattggatttc	34d
Db	646	TTATGTTTCAAGACATGCAGGAAAATATGATTCCAGATAAATCATAAAAATTCAAATTTAAAA	705
Qy	346	tcagcaactctaagaattgatagacaaaaggaatttagtatataaacacaagagaa	405
Db	706	TATCCAGCTGTTTATGATGACAAGATAAAAAAGTGTCATATATTATATATTCGAGCTCAA	765
Qy	406	agtaataa 413	
Db	766	GAAAATAA 773	
RESULT 8			
US-08-257-073-8			
:	:	: Sequence 8, Application US/08257073	
:	:	: Patent No. 5766597	
:	:	: GENERAL INFORMATION:	
:	:	: APPLICANT: Paoletti, Enzo	
:	:	: APPLICANT: de Taisne, Charles	
:	:	: APPLICANT: Tine, John A.	
:	:	: TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE	
:	:	: NUMBER OF SEQUENCES: 143	
:	:	: CORRESPONDENCE ADDRESS:	
:	:	: ADDRESSEE: Curtis, Morris & Safford, P.C.	
:	:	: STREET: 530 Fifth Avenue, 25th Floor	
:	:	: CITY: New York	
:	:	: STATE: New York	
:	:	: COUNTRY: UNITED STATES OF AMERICA	
:	:	: ZIP: 10036	
:	:	: COMPUTER READABLE FORM:	
:	:	: MEDIUM TYPE: Floppy disk	
:	:	: COMPUTER: IBM PC compatible	
:	:	: OPERATING SYSTEM: PC-DOS/MS-DOS	
:	:	: SOFTWARE: PatentIn Release #1.0, Version #1.30	
:	:	: CURRENT APPLICATION DATA:	
:	:	: APPLICATION NUMBER: US/08/257,073	
:	:	: FILING DATE: 09-JUN-1994	
:	:	: CLASSIFICATION: 424	
:	:	: PRIOR APPLICATION DATA:	
:	:	: APPLICATION NUMBER: US 08/075,783	
:	:	: FILING DATE: 11-JUN-1993	
:	:	: PRIOR APPLICATION DATA:	
:	:	: APPLICATION NUMBER: US 07/852,305	
:	:	: FILING DATE: 18-MAR-1992	
:	:	: PRIOR APPLICATION DATA:	
:	:	: APPLICATION NUMBER: US 07/672,183	
:	:	: FILING DATE: 20-MAR-1991	
:	:	: ATTORNEY/AGENT INFORMATION:	
:	:	: NAME: Frommer, William S.	
:	:	: REGISTRATION NUMBER: 25,506	
:	:	: REFERENCE/DOCKET NUMBER: 454310-2570	
:	:	: TELECOMMUNICATION INFORMATION:	
:	:	: TELEPHONE: (212) 840-3333	
:	:	: TELEFAX: (212) 840-0712	
:	:	: TELEX: 425066 CURTMS	
:	:	: INFORMATION FOR SEQ ID NO: 8:	
:	:	: SEQUENCE CHARACTERISTICS:	
:	:	: LENGTH: 1884 base pairs	
:	:	: TYPE: nucleic acid	
:	:	: STRANDEDNESS: single	
:	:	: TOPOLOGY: linear	
:	:	: US-08-257-073-8	

```
Query Match          5.2%   Score 48.8; DB 1; Length 1884;
Best Local Similarity 47.4%; Pred. No. 0.024;
Matches 146; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
```

Qy	166	gtagataatagaataaaaactcttgctaataatgaagaaaaatttggtttagcatcatcaagtgttg	228
Dd	526	TTAAGATGGAGGTTTGTCTTCCTCAACAGAACCTCTTATGTGCACCAATGACATTGA	585
Qy	226	aaceaaacggtgaaaatttcctcaggagaatatattttitagaatggatgcgatgatatt	285
Dd	586	GATGAATGAGACATTTTTATAAGAGATAAATAATGTAAAAAANTTTAGATGAATTGACT	645
Qy	286	tcatatccaaagtagatttgtatagaagaaaactcgtttttatggaggaaaaatcattggatttc	345
Dd	646	TTATGTTCAAGACATCAGGAATATGATTCAGATAATGATNAANAATTCANATTTATAA	705
Qy	346	tcagcaactctaagaattgtagaccacaaaaggaaatttagtatataaacacaagagaa	405
Dd	706	TATCCAGCTGTTTATGATGACAAGATAAAAAGTGTCATATATTATATATTTGCAGCTCAA	765
Qy	406	agtaataa 413	
Dd	766	GAATAAA 773	
RESULT 9			
US-08-998-416-186/C			
; Sequence 186, Application US/08998416			
; Patent No. 6239264			
; GENERAL INFORMATION:			
; APPLICANT: Philippsen, Peter			
; APPLICANT: Pohlmann, Rainer			
; APPLICANT: Steiner, Sabine			
; APPLICANT: Mohr, Christine			
; APPLICANT: Wendland, Jurgen			
; APPLICANT: Knechtle, Philipp			
; APPLICANT: Rebschung, Corinne			
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII			
; TITLE OF INVENTION: AND USES THEREOF			
; NUMBER OF SEQUENCES: 1152			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: No. 6239264artis Corporation			
; STREET: 3054 Cornwallis Road			
; CITY: Research Triangle Park			
; STATE: No. 6239264th Carolina			
; COUNTRY: USA			
; ZIP: 27709			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/998,416			
; FILING DATE: 24-DEC-1997			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: CH 0016/97			
; FILING DATE: 31-DEC-1996			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Meigs, J. Timothy			
; REGISTRATION NUMBER: 38,241			
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGCI976			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 919-541-8587			
; TELEFAX: 919-541-8689			
; INFORMATION FOR SEQ ID NO: 186:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 615 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: DNA (genomic)			
; ORIGINAL SOURCE:			
; ORGANISM: PAG1074RP			
US-08-998-416-186			

```
Query Match      5.2%; Score 48.6; DB 4; Length 615;
Best Local Similarity 49.2%; Pred. No. 0.022;
Matches 157; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY 576 ttccaaatcggttactcaataaagctactcttttacagtatagattaaacagagaatggaat 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 TTATAATATTATTATTAATAATTAATAATGATAAATAATTAATAA---ATAATTAATT 552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 636 atcacaaacaaataagtttaagcaataattttactcagctattttcaagaatttttataa 695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 TAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 696 agaaaaatctatttgatcacaaaaattactaatcttctcaagatgtagataaa 755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 ATAGAATAATTAAGCTTAATAATTAATAATAATAATAATAATAATAATAATAATA 432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 756 gaaacgtctactcagcagcgtctctctaaatttttgagctaaattctaccctagat 815
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 AAATCAACATAAATTTATATAAATAATAGATATATATAATAAATAATAATAATAATTT 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 816 tactattagaaactatatattgtttatttatatttactttaagctcccttggttagag 875
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 AAATAAATAAATCTTATATAATAATAATAATAATAATAATAATAATAATAATAATAA 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 876 gttattaataaatgatatt 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 TATATTAAATTATGATAAT 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-961-083-195
; Sequence 195, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 195:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 844 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-195
```

```
Query Match      5.1%; Score 47.6; DB 3; Length 844;
Best Local Similarity 50.4%; Pred. No. 0.039;
Matches 116; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 103 ataattgtcattgataatcccaagtagagtgatttaaagcaattcttaacagaataattca 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 ATTATTTTAGTTGATGACGGTTCTACGGATAATTCTCGGGAAATTTGTGATGCTTTTATG 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 163 gtttagataataaataaaatcttcttaataagaagaaatattggttttagcatcaagt 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 ATGCAAGATAATCGTGCAGATATTGTCATCAAGAAATAAGGGGGGCGACACAAGCT 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 223 ttgacaagaacggtgaaatcttctaaaggaggaataatatatttagaattgagtgctgat 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 AAAATATGGGATAGTAGTGACCTAAGGAGAGTACATCAGCATTTGTTGATTCAGATGAT 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 283 atttcataatcccaagtagatttgataagcaaatcttctttagggagaaaa 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 ATCGTAAAGAAATAATGATTGAAACTCTTTATCAGCAAGTCCCAAGAAAA 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; NUMBER OF SEQUENCES: 1152
; NUMBER OF INVENTION: AND USES THEREOF
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
; US-08-998-416-288
```


TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-990-571-10

Query Match 4.9%; Score 46; DB 4; Length 3701;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 151; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```
QY 593 tcaataaagctactttacagtatagattaaacgagaatggaatatcacaaaccaataagt 652
DB 581 TAATTATAATCAATTATTATATATATATATATGGATATGATATATATATATATATTAATGATA 522
QY 653 ttaagcaatatatttactcagctattttacaagattttttataaagaaaaatccttatattg 712
DB 521 CAATAACAATATGTAATTATATAATAATCAATAATAATCATATATATAATAAATTAATATA 462
QY 713 atatacaaaaaattactaattactttcaagagatgtgataaagaacgcctatactcagc 772
DB 461 ATATTGCAATTATTAATTATATATATCTGTTCTTAATTTATATCATTAACAATTACAAAACATA 402
QY 773 aagagctctctaaatatatttgagctaaatctacccttagtattactattagaaaaactat 832
DB 401 TATACITTTTATAATTATAACACATGAATATATTAATATATATTAACAATGCCTATG 342
QY 833 atatttggttatttatactttaagtcctcccttggttagaggttattataataaagata 892
DB 341 ATGATTATTATCATCTTAAATTCGTCAATTAATCCCTTAAGATCTTATCTAATTTATTTAT 282
QY 893 ttaatatatttagtactgaaattggtt 918
DB 281 TTAATAATTATCTATACTGATATATTTT 256
```

RESULT 14

US-08-723-142A-10/G
; Sequence 10, Application US/08723142A
; Patent No. 6306396
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,142A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-723-142A-10

Query Match 4.9%; Score 46; DB 4; Length 3701;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 151; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```
QY 593 tcaataaagctactttacagtatagattaaacgagaatggaatatcacaaaccaataagt 652
DB 581 TAATTATAATCAATTATTATATATATATATGGATATGATATATATATATATTAATGATA 522
QY 653 ttaagcaatatatttactcagctattttacaagattttttataaagaaaaatccttatattg 712
DB 521 CAATAACAATATGTAATTATATAATAATCAATAATAATCATATATATAATAAATTAATATA 462
QY 713 atatacaaaaaattactaattactttcaagagatgtgataaagaacgcctatactcagc 772
DB 461 ATATTGCAATTATTAATTATATATCTGTTCTTAATTTATATCATTAACAATTACAAAACATA 402
QY 773 aagagctctctaaatatatttgagctaaatctacccttagtattactattagaaaaactat 832
DB 401 TATACITTTTATAATTATAACACATGAATATATTAATATATATTAACAATGCCTATG 342
QY 833 atatttggttatttatactttaagtcctcccttggttagaggttattataataaagata 892
DB 341 ATGATTATTATCATCTTAAATTCGTCAATTAATCCCTTAAGATCTTATCTAATTTATTTAT 282
QY 893 ttaatatatttagtactgaaattggtt 918
DB 281 TTAATAATTATCTATACTGATATATTTT 256
```

RESULT 15

US-08-998-416-1137/C
; Sequence 1137, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steinet, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtie, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NO. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1692RP
;
US-08-998-416-1137

```

```

Query Match      4.9%; Score 45.6; DB 4; Length 636;
Best Local Similarity 50.6%; Pred. No. 0.099;
Matches 162; Conservative 0; Mismatches 154; Indels 4; Gaps 2;

QY 576 ttctaaatcggcttactcaataaggtacttttacagtgatagattaaacga-gaatggaa 634
    || || || || || || || || || || || || || || || || || || || || ||
Db 609 TTATAATATTTTATTAAATTAATTAATAATGATAAAATATTAAATAATTAATTAATTTAA 550

QY 635 tatcacaacccaataagtttaagcaatatatttactcagctattttacaaagattttta 694
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 549 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 490

QY 695 agaaaaatcttatattgtatcacaaaaattactaattacttccaagagtgatgataa 754
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 489 AGAAATTAAAGTTAAATTAATTTTAAATAATAATCTTTATAAAAGATTAAATAATAATA 430

QY 755 agaaacgctatactcagcagagctctcctaaatattttgagctaaatactacccctagta 814
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 429 ATCAACATAATAATTTTATAAAATAGATAT--TATAATAAAATAATAATAATTTACATAAT 373

QY 815 ttactattagaaactatatttggtttatatttatacttttaagctccctggtagga 874
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 372 TAAATAAATAAATCCTTTATAATAATAATAATAATAATAATAATAATAATAATAATA 313

QY 875 ggttattaataaatgatatt 894
    ||||| | | | | |
Db 312 ATATATTAATTATTGATAAT 293

```

Search completed: August 2, 2002, 13:28:36
Job time: 9682 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 18:56:25 ; Search time 13.52 Seconds
(without alignments)
896.392 Million cell updates/sec

Title: US-09-900-038A-1

Perfect score: 1590

Sequence: 1 MNYGIIMSVYNEPLNYRDS.....LINDINILVLKFGKQSD 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	437.5	27.5	290	YA57_METJA	Q58457 methanococ
2	284	16.6	323	YF78_HAEIN	Q57287 haemophilus
3	206.5	13.0	267	YG95_HAEIN	Q48215 haemophilus
4	199.5	12.5	266	AMSE_ERWAM	Q46635 erwania amy
5	196.5	12.4	294	YG96_HAEIN	Q48214 haemophilus
6	192	12.1	900	GGAB_BACSU	Q46918 bacillus su
7	189	11.9	250	Y868_HAEIN	Q57022 haemophilus
8	182.5	11.5	344	YIBD_ECOLI	P11290 escherichia
9	175	11.0	256	SPSA_BACSU	P39621 bacillus su
10	161	10.1	279	WCAA_ECOLI	P77414 escherichia
11	158.5	10.0	909	Y4GI_RHISN	P55465 rhizobium s
12	145	9.1	268	YWDF_BACSU	P39614 bacillus su
13	144.5	9.1	322	YS86_ANASP	P22639 anabaena sp
14	140	8.8	323	YKCC_BACSU	Q34319 bacillus su
15	129	8.1	301	AMSB_ERWAM	Q46632 erwania amy
16	125	7.9	260	DPML_MOUSE	O70152 mus musculus
17	123	7.7	266	DPML_CRIGR	Q9WU83 cricetulus
18	122.5	7.7	346	YF20_MYCTU	Q50587 mycobacteri
19	119	7.5	260	DPML_HUMAN	O60762 homo sapien
20	119	7.5	581	Y208_BORBU	O51226 borrelia bu
21	119	7.5	1197	DPOM_PODAN	Q01529 podospora a
22	118	7.4	446	GGAA_MYCSU	P46917 bacillus su
23	116.5	7.3	979	P115_MYCHR	P41508 mycoplasma
24	116	7.3	1805	HMW2_MYCGE	P47460 mycoplasma
25	114.5	7.2	236	DPML_SCHPO	O14466 schizosacch
26	113.5	7.1	993	NISE_LACLA	P21003 lactococcus
27	113	7.1	569	CYSP_PLAFA	P25805 plasmodium
28	112	7.0	419	HASA_STRPY	Q54865 streptococc
29	110.5	6.9	997	T257_ECOLI	P25239 escherichia
30	110	6.9	2273	HFAL_YEAST	P32874 saccharomyc
31	109.5	6.9	653	MTSL_STRSA	P29347 streptococc
32	109	6.9	970	Y087_BUCAI	P57189 buchnera ap
33	107	6.7	1169	EX5B_BORBU	O51578 borrelia bu

34	106.5	6.7	299	1	Y060_MYCPN	P75042 mycoplasma
35	106	6.7	342	1	EXOU_RHIME	P33700 rhizobium m
36	106	6.7	1070	1	EX5C_BUCAI	P57528 buchnera ap
37	104.5	6.6	329	1	CSBB_BACSU	O45539 bacillus su
38	104.5	6.6	1104	1	SYV_YEAST	P07806 saccharomyc
39	104	6.5	426	1	CGED_BACSU	P42092 bacillus su
40	104	6.5	582	1	BPS2_ACIAM	P32985 acidianus a
41	104	6.5	610	1	MUTL_BORBU	O51229 borrelia bu
42	104	6.5	1081	1	PDR6_YEAST	P32767 saccharomyc
43	103.5	6.5	1786	1	YCF1_ARATH	P56785 arabidopsis
44	102.5	6.4	319	1	YF18_MYCTU	Q50590 mycobacteri
45	102.5	6.4	1275	1	RFBC_MYXXA	Q50864 myxococcus

ALIGNMENTS

RESULT 1
YA57_METJA STANDARD; PRT; 290 AA.
AC Q58457;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase MJ1057 (BC 2.-.-.-).
GN MJ1057.

OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.";
RL Science 273:1058-1073(1996).

CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U67549; AAB99061.1; -.
DR HSSP: P39621; 1QG8.

DR TIGR: MJ1057; -.

DR InterPro: IPR001173; Glycos.transf_2.

DR Pfam: PF00535; Glycos.Transf_2; 1.

KW Hypothetical protein; Transferase; Complete proteome.

SQ SEQUENCE 290 AA; 35099 MW; 3F6A1B221C420D74 CRC64;

Query Match 27.5%; Score 437.5; DB 1; Length 290;

Best Local Similarity 35.2%; Pred. No. 1.7e-22;

Matches 99; Conservative 58; Mismatches 101; Indels 23; Gaps 8;

QY 4 SLIMSVYNEPLNYRDSVESILNQTLDTFEPIIVDNPGRDLKQFLTEYSWPNRIKIL 63

Db 11 SVVWATNEFEKYUKESIESIXNXTKDFEPIIVLDNPNKKAEITEKYQOKRIRIFI 70

QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN-SLDFSATLIEL 122

```
Db 71 KNERLGRGASRKNKAVNIARGKYIAILDADDIALPKRLEKQFKYMENRDRIDLLESWYF 130
QY 123 IDQGNLVYKQESNKIYLTNDIRKMLNRSILAHPTWCVRKKVFDKLMGYRDLVPVEDY 182
Db 131 IDENGIL-KEPKPK-KYFREIKKYFKEHLTVHPMMVSKILKLKYDEKLRSQDY 188
QY 183 DPAIGALADFKIGLNLKVLQYRL--NENGISQTNKFKQYIYSAILQDFYKEKSY---- 236
Db 189 DFWIRCIANDYKFDIEFLLKYRIPNRDNYLSRIKKQKLYSYT-LKTHWKNKKHFCNN 247
QY 237 IDITKITNYFOEYVVKKRYTQOELSYPKLSKTPSITIRKL 277
Db 248 VYFVKV--FFYSLVV-----YLFVLVTPFTILKIL 275

RESULT 2
YF78_HAEIN STANDARD; PRT; 323 AA.
AC Q57287; O05077;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase HI1578 (EC 2.-.-.-).
GN HI1578.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.;"
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32832; AAC23227.1; -.
DR TIGR; HI1578; -.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2.1.
KW Hypothetical protein; transferase; Glycosyltransferase;
SQ SEQUENCE 323 AA; 37680 MW; 7CBC2681039AB5B4 CRC64;

Query Match 16.6%; Score 264; DB 1; Length 323;
Best Local Similarity 29.0%; Pred. No. 6.5e-11;
Matches 93; Conservative 62; Mismatches 114; Indels 52; Gaps 13;

QY 4 STMSVYNEPLNVYRDSVESILNQTLTFEFTIIVDNPGRGDLKQFLTEYSVVDNRKIL 63
Db 8 SVIVCAYNAE-QYIDESISSLIQNYENLE-IVINDGSTDLTSLHLEISIKLDKRKII 65
QY 64 LNEENIGLASSLUNKAVKISGEYIFRMDADDISYSPRFDKQIRFMEENS-LDFSATLIEL 122
```

```
Db 66 SNKYNLGFSINLIGLGCFSKGYFARMDDADDIAKPSWIEKIVTYLEKNDHITAMGSYLEI 125
QY 123 IDQGNLVYKQESNKIY----LTNDIRKMLNRSILAHPTWCVRKKVFDKLMGYRDLVPVEDY 175
Db 126 IVEKCGIIGSOGYKTDIWKNPPLHNDICEAMLFNPHNTNMTIMRANVYRHKILFNKD 185
QY 176 LVPVEDYDF----AIRGALADFKIGLNLKVLQYRLNENGISQTNKFKQYIYSAILQ---- 228
Db 186 YPAEDYKFWSEVSRGLGLANYP-----EALVKYRLHGNOTSSVYNHEQNETAKKIKREN 240
QY 229 -DFYKEKSYDITKITN-----YFOEYVVKKRYTQOEL---SKY--- 263
Db 241 ITYLNKIGIDIKVINSVLSLEIYHVDKSNKVLKSLIYEMYSLDKYTTITSLHFIKYHL 300
QY 264 --FELKSTPSIT---IRKLYI 279
Db 301 ELFDUKQNLKIIRKINV 321

RESULT 3
YG95_HAEIN STANDARD; PRT; 267 AA.
AC Q48215; O05081;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase HI1695 (EC 2.-.-.-).
GN HI1695.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A2;
RA McLaughlin R., Abu Kwaik Y., Young R., Spinola S., Apicella M.;
RT "Characterization and sequence of the lsg locus from Haemophilus
RT influenzae.;"
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.;"
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M94855; AAA24983.1; -.
DR EMBL; U32842; AAC23341.1; -.
DR TIGR; HI1695; -.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2.1.
KW Hypothetical protein; transferase; Glycosyltransferase;
KW Complete proteome.
```


FT CONFLICT 26 26 V -> G (IN REF. 1).
 FT CONFLICT 46 46 D -> E (IN REF. 1).
 FT CONFLICT 49 49 F -> S (IN REF. 1).
 SQ SEQUENCE 267 AA; 30770 MW; A2F1A0532737D8C3 CRC64;

Query Match
 Best Local Similarity 13.0%; Score 206.5; DB 1; Length 267;
 Matches 70; Conservative 57; Mismatches 130; Indels 25; Gaps 8;

QY 1 MNYSIMSVY-NEPLNVRDVSVESILNQTLDTEFEIIVIDNPSRGDLKQFLTEYSVVDNR 59
 Db 1 MKFSVLSLKIYKENPQELFCFESLVAQTRQADEIIVLFDGVVTPDLEFVYTFEE-TKLP 59
 QY 60 IKILLNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN--SLDPSA 117
 Db 60 LKLVLPQNLGLKALNEGILLHCDYDWFVMDTDDICVDFEQAQVAFIQHPESIIFGG 119
 QY 118 TLIELIDOKGNLV-YKQRESNKIYLTNDIRKMLNRSILAHPTWCVKKKVFDKLMGYRDL 176
 Db 120 QIAEFGRNVDIVAYRNVPTS---AQELIKFTQKRCPPFNHMTVAYQKSAVINGCGYEDL 175
 QY 177 VPVEDYDFAIRGALADFKIGLLNKVLQYRLNENGISQTNKFKQYIYSAILQDFYKESY 236
 Db 176 --QEDYVLTWIKLVAQGLYMANLPDILYVYRNGVMWSRRRGVNO---AKAEWRLFKLYR 230
 QY 237 IDITKITNYFOEYVKKRYTQQLSKYFELKSTPSIIRKLY 278
 Db 231 LGTQGLSLGFTFALR-----FGSRLPTSLKKLY 261

RESULT 4
 AMSE_ERWAM STANDARD; PRT; 266 AA.
 AC Q4635;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Amylovoran biosynthesis glycosyl transferase amse (EC 2.-.-.-).
 GN AMSE.
 OS Erwinia amylovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwinia.
 OX NCBI_TaxID=552;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EAL/79;
 RX MEDLINE=95319333; PubMed=7596293;
 RA Bugert P., Geider K.;
 RT "Molecular analysis of the amse operon required for exopolysaccharide
 synthesis of Erwinia amylovora.";
 RL Mol. Microbiol. 15:917-933(1995).
 CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVORAN WHICH
 CC FUNCTIONS AS A VIRULENCE FACTOR.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; X77921; CAA54886.1; -;
 DR InterPro; IPR001173; Glycos_transf_2.
 DR Pfam; PF00535; Glycos_transf_2; 1.
 KW Exopolysaccharide synthesis; Transferase; Glycosyltransferase.
 SQ SEQUENCE 266 AA; 30748 MW; 163268A4210EB47B CRC64;

Query Match
 12.5%; Score 199.5; DB 1; Length 266;

Best Local Similarity 28.4%; Pred. No. 9.9e-07;
 Matches 73; Conservative 48; Mismatches 107; Indels 29; Gaps 12;

QY 3 YSIIMSVMY-NEPLNVRDVSVESILNQTLDTEFEIIVIDNPSRGDLKQFLTEYSVVDNR 60
 Db 2 FSVLISYNKEKPN-LEQCLESJHQOTLNADEIIVLVDGVPVSESLKAVATRW----NLL 57
 QY 61 KILL-NEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN-SLDPSA 117
 Db 58 PLVIVPLEKNLGLKALNAGLERCTHNVVARMDDTDDICLPERFEKQISYMESHPEVVLGS 117
 QY 118 TLIELIDOKGNLV-YKQRESNKIYLTNDIRKMLNRSILAHPTWCVKKKVFDKLMGYRDL 172
 Db 118 AAVIEFDEHG---KERLRLPLSNNDIHEFARMKNPNH--MCV---VFRKDKVISAGS 168
 QY 173 YRLDVPVEDYDFAIRGALADFKIGLLNKVLQYRLNENGISQTNKFKQYIYSAILQDFYK 232
 Db 169 YQHLYLMEDYNLWLRIMSLGHPVANLPDVLKVR----GSDMVNKRNGWNIKSEVQLYR 225
 QY 233 EKSVIDIT---KITNYF 246
 Db 226 LKALKQTGFIRGTLF 242

RESULT 5
 YG96_HAEIN STANDARD; PRT; 294 AA.
 AC Q48214; Q05082;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Putative glycosyl transferase H11696 (EC 2.-.-.-).
 GN H11696.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A2;
 RA McLaughlin R., Abu Kwaik Y., Young R., Spinola S., Apicella M.;
 RT "Characterization and sequence of the isg locus from Haemophilus
 RT influenzae.";
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; M94855; AAA24982.1; -;
 DR EMBL; U32842; AAC23342.1; -;
 DR TIGR; H11696; -;

DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
FT Complete proteome.
KW CONFLICT 38 C -> Y (IN REF. 1).
FT CONFLICT 48 S -> R (IN REF. 1).
FT CONFLICT 70 V -> I (IN REF. 1).
FT CONFLICT 74 T -> A (IN REF. 1).
FT CONFLICT 88 C -> R (IN REF. 1).
FT CONFLICT 97 V -> I (IN REF. 1).
FT CONFLICT 106 N -> D (IN REF. 1).
FT CONFLICT 152 I -> V (IN REF. 1).
FT CONFLICT 185 P -> S (IN REF. 1).
SQ SEQUENCE 294 AA; 33646 MW; 8330F081BFF7A18B CRC64;

Query Match 12.4%; Score 196.5; DB 1; Length 294;
Best Local Similarity 25.1%; Pred. No. 1.8e-06;
Matches 81; Conservative 60; Mismatches 115; Indels 67; Gaps 15;
QY 4 SLIMSVYNEPLNLYVRSVESILNQTLDTEFIIVDNPGRGDLKQFLTEYSVVDNR--- 59
DB 3 STIIVSYNRKAE-VPALLESLTQQTSSNFE-VLIIVDCSK-----ERVVVEQSYSPF 52
QY 60 IKILNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN---SLDPS 116
DB 53 VTVIRNETNOGAESRNVGARTSKGDLWLLFLDDDDCFWPKCEKVLQVTEQNPINFIYH 112
QY 117 APLIELDQKGNLVYKQESNKIYLTNDIRKMLNRSILAHPTWCVKKVDKLMGYRD- 175
DB 113 PAKCEWNEGFTYVQPIEQEI-----STERILLANKGGMPWIAIKEMFLKIGGLSTA 168
QY 176 LVPVEDYDFAIRGALADFKIGLLNKVLQ-----YRLNE-----NGISQ 214
DB 169 LRSLEDYDFLLK-----LLQEPSFTYKINEPLTYCTFHTKRSVSDTDTN 214
QY 215 TNKFQYIYSAILQDPYEKSYVIDTKTNTYQEVYIKRKYTOQLSKYFEL-KSTPSIT 273
DB 215 TOKAIDYIRHVVYKVEQARNF-DIN--ASYILAYPHIMNLNRSKAAKYVDFIKTKSI- 270
QY 274 IRKLYICLYLPKSPILVRLLIN 296
DB 271 --KQFIITLVILISP---KLAIN 288

RESULT 6
GGAB_BACSU STANDARD; PRT; 900 AA.
AC P46918;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Minor teichoic acids biosynthesis protein ggab.
GN GGAB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Freymond P., Karamata D.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GALACTOSAMINE-CONTAINING
CC MINOR TEICHOIC ACIDS.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
DR EMBL: U13979; AAA73513.1; -.
DR EMBL: Z99122; CAB15585.1; -.
DR Subtilist; BG11192; ggab.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 900 AA; 107154 MW; FA6649548C2C62F CRC64;

Query Match 12.1%; Score 192; DB 1; Length 900;
Best Local Similarity 23.7%; Pred. No. 1.4e-05;
Matches 88; Conservative 69; Mismatches 130; Indels 84; Gaps 19;

QY 2 NYSIIMSVYNEPLNLYVRSVESILNQTLDTEFIIVI-----DNPSRGDL--KQFLTEYSV 55
DB 9 DFSVIMPIYNVEL-YLTEAIESIINQTI-GFENIQILVNDSDSPKSEIICKEYAQY-- 64
QY 56 VDNRIKILLNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEENSLDF 115
DB 65 -PNNI-VYAKKQNGGVSSARNYGLXVAGRYIQFLDPPDLVSECTFENVLNFDEHKNEI 122
QY 116 SATLIELDQKGNLVYKQESNKIYLTN-DIRKMLNRSILAHPTWC-----VKKKVFQK 169
DB 123 DIVAIPFIFFAAGR-TGEHNLNKFSTRILDVEKE--PHILTH---CGSTFIKKDALKN 176
QY 170 LMGYRDLVPVEDYDFAIRGALADFKIGLLNKVLQVRLNENGIS--QTNKFQYIYSAIL 227
DB 177 IRFDENCKIGEDAKLVNLIISOKKYGVLVKEAKYHYVRDGGSSAMQATAKANNMFNHL 236
QY 228 QDFYKESYIDITKITNY-----FQEYVI-----KKRYTQ----- 257
DB 237 IFF--SKNLIDIIK--NHEQKIPFLQYVMHDLKWLKLIKIDSETPLDENYSEPLTLI 292
QY 258 QELSXYFE---LKSTPSITIRKLYICL-----YLYFKSPILVRRL 293
DB 293 REVLSYIDDDVLIETKSVSHFYHALKIKHGENYSRYVYERETEQDYLYLREGKIVSKL 352
QY 294 LINDINILVLK 304
DB 353 SDQTLTIEILE 363

RESULT 7
Y868_HAEIN STANDARD; PRT; 250 AA.
AC Q57022; P96336;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase HI0868 (EC 2.-.-.-).
GN HI0868
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kesteven A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd".
RL Science 269:496-512(1995).


```
Db 123 DYVQ--GEVYSOPASLPYKPSYRRLFYKRNIIGNOVFTWAWRFEKCLFD-----TE 175
Qy 176 LVPVEDYDFAIRGAL---ADPKILLNKLVLQYRLNENGI---SOTNKEFOYIYSAILQD 229
Db 176 LKAQDIDIFLRVVEYGEWPKVVEATQIL---HINIGEMQITSSPKKFSGYFH----- 226
Qy 230 FY-KESYID-ITKITNYFOEYVIK-KRYTQOEL 260
Db 227 FYRKHKKDFDRASKYQFLTLYQIRNKRMTWRTL 260

RESULT 11
Y4GI_RHISN
ID Y4GI_RHISN STANDARD; PRT; 909 AA.
AC P55465;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 102.8 kDa protein Y4GI.
GN Y4GI.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: STRONG TO M.XANTHUS O-ANTIGEN BIOSYNTHESIS PROTEIN
CC REBC (NOT TO BE CONFUSED WITH ENTEROBACTERIAL REBC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000074; AAB91683.1; -
DR HSSP; P39621; 1QG8.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 2.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 909 AA; 102827 MW; 3985D69722F43962 CRC64;

Query Match 10.0%; Score 158.5; DB 1; Length 909;
Best Local Similarity 22.6%; Pred. No. 0.0023;
Matches 53; Conservative 51; Mismatches 102; Indels 29; Gaps 6;

Qy 4 SIIMSVYNEPLNYVRDVSVESILNQTLDPEFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db LVVVVYVNPDPALLVEMESVRAQSYANWELCLDDCSTDPEVGRVLRNRYAAQDPVRVV 428
Qy 64 LNEENIGLASSLNKAVKISKEYIFRMDADDISYPSRFDKQIRWEENSLDFAATLIELI 123
Db 429 FREANGHMSQASNAIEIARGAYIALDHDLDLP-----DALVLVQVVI 473
Qy 124 DQGNLYVKQRESNKIYLTNDIR-----KMLNRSILAHPTWCVKKKVFDKLM-----GY 173
Db 474 DAHPDAKIITYDEDKI-VEGTRCDIAHKPDWNRDLVYINGIISHLGVDFDAALVREVGAF 532
Qy 174 RD-LVPVEDYDFAIR--GALADFKIGLLNKVLQYRLNENGISOTNKFQKQIYSA 225
Db 533 REGFEGAGDYDMLLRCTIERVQDRQIHIIHIAKVLYSWRATPGSAAASNRAPKPYANEA 587

RESULT 12
```

```
YWDF_BACSU
ID YWDF_BACSU STANDARD; PRT; 268 AA.
AC P39614;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase ywdf (EC 2.-.-.-).
GN YWDF OR IPA-56D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73124; CAAS1612.1; -
DR EMBL; Z99123; CAB15824.1; -
DR PIR; S39711; S39711.
DR Subtilist; BGI0602; ywdf.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 268 AA; 30616 MW; DD8428F7016EC9B3 CRC64;

Query Match 9.1%; Score 145; DB 1; Length 268;
Best Local Similarity 25.1%; Pred. No. 0.0042;
Matches 68; Conservative 50; Mismatches 125; Indels 28; Gaps 11;

Qy 1 MNYSIIMSVYNEPLNYVRDVSVESILNQTLDPEFIIVID-NPSRGDLKQFLTEYSVVDNR 59
Db 1 MKISIVIVTYNR-IPALCELLESISRQTLMPYEIIIVNDAGESVVPVKALYPELP----- 54
Qy 60 IKILLNENIGLASSLNKAVKISKEYIFRMDADDISYPSRFDKQIRWEENSLDFAATL 119
Db 55 IAVINLEKNSGHVAAARNAGVKEASGDCIMLCDDDDFFTPPGHIEKMAKETIADTFVHSDAE 114
Qy 120 IELDOKGNLYVKQRESNKIY-LTNDIRKMLNRSILAHPTWCVKKKVFDKLMGY--RD 176
Db 115 IVSPEEKNGTYR--VSRKLFAYTADYEDMRVFSYV--PGSWMYRRLHDEIGYFDADV 170
Qy 177 VPEVDYDFAIRGALADFKIGLLNKVLQYRLNENGISOTNKFQKQIYSAILODFYK 232
Db 171 HNYWDWDFYLRAA-KDYRVKRVPCASVIYAFSDAGDQNSADLGAKRKOYL-----DRLS 223
Qy 233 EKSVIDITKIYNY----FOEYVIKKRYTQOEL 260
Db 224 EKHLGELPTKFNFAVLLEPEMKRREKSEM 254

RESULT 13
YS86_ANASP
ID YS86_ANASP STANDARD; PRT; 322 AA.
AC P22639;
```

[illegible]

2

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 18:51:04 ; Search time 17.53 seconds
(without alignments)
1715.685 Million cell updates/sec

Title: US-09-900-038A-1
Perfect score: 1590
Sequence: 1 MNYSIMSYNEPLNVRDS.....LINDINILVLKFGGKQSD 313

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	431.5	27.1	290	2 H64431	glycosyl transfera
2	367	23.1	333	2 H97167	glycosyltransferas
3	348.5	21.9	336	2 A97168	glycosyltransferas
4	317	19.9	333	2 B97168	glycosyltransferas
5	300.5	18.9	271	2 B84114	exopolysaccharide
6	282	17.7	340	2 T44330	glycosyl transfera
7	271	17.0	367	2 G95948	probable glycosyl
8	264	16.6	323	1 H64130	glycosyl transfera
9	251	15.8	321	2 AG2188	hypothetical prote
10	246.5	15.5	257	2 E84107	telchuronic acid b
11	221.5	13.9	329	2 AB2092	glycosyltransferas
12	221	13.9	278	2 D70036	exopolysaccharide
13	220	13.8	252	2 A69728	biosynthesis of te
14	220	13.8	604	2 E97757	hypothetical prote
15	218.5	13.7	333	2 AH2026	hypothetical prote
16	218	13.7	260	2 E90984	probable glycosyl
17	218	13.7	260	2 H85829	glycosyl transfera
18	217.5	13.7	280	2 E71703	glycosyl transfera
19	216.5	13.6	281	2 G97777	glycosyl transfera
20	215.5	13.6	318	1 E71690	minor telchoic aci
21	215.5	13.6	346	2 H81970	lacto-N-neotetraos
22	212.5	13.4	343	2 AI2091	glycosyltransferas
23	210.5	13.2	318	2 AG2189	hypothetical prote
24	208	13.1	294	2 E83022	probable glycosyl
25	206.5	13.0	267	2 C64175	hypothetical prote
26	206.5	13.0	294	2 G71148	probable glycosyl
27	205.5	12.9	334	1 G71153	hypothetical prote
28	205	12.9	349	2 D81027	lacto-N-neotetraos
29	204.5	12.9	291	2 F95015	glycosyl transfera

glycosyl transfera
glycosyl transfera
glycosyl transfera
capsular polysacch
spore coat polysac
glycosyl transfera
galactosamine-cont
glycosyltransferase
hypothetical prote
glycosyltransferas
glycosyl transfera
beta-1,4-galactosy
hypothetical prote
probable glucosyl
hypothetical prote

30 203 12.8 333 1 S70813
31 202.5 12.7 317 2 H97888
32 202 12.7 268 2 B97189
33 201 12.6 344 2 A70037
34 198 12.5 298 2 B75096
35 197 12.4 344 2 G70036
36 196.5 12.4 294 2 D64175
37 192 12.1 900 2 E89631
38 191.5 12.0 258 2 S61898
39 190.5 12.0 318 2 AH2189
40 190.5 12.0 465 2 T00090
41 189 11.9 250 2 A64099
42 188 11.8 318 2 T50039
43 187.5 11.8 596 2 AD1876
44 186 11.7 274 2 E81319
45 185.5 11.7 996 2 AG2366

ALIGNMENTS

RESULT 1

H64431

glycosyl transferase (EC 2.4.-.-) - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997

C:Accession: H64431

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: H64431

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-290 <BUIL>

A:Cross-references: GB:U67549; GB:L77117; NID:g1591709; PID:g1592318; TIGR:MJ1057; PI

C:Genetics:

A:Map position: FOR996513-997385

C:Keywords: glycosyltransferase

Query Match 27.1%; Score 431.5; DB 2: Length 290;
Best Local Similarity 34.9%; Pred No. 9.7e-21;
Matches 98; Conservative 58; Mismatches 102; Indels 23; Gaps 8;

Qy 4 SIIMSVYNEPLNVRDSVESILNQTLDFEFIVIDNPSRGDLKQFLTEYSYVDNRKIL 63

Db 11 SVMATYNEPEKYLKESIESIXNQTKDQFXFIIXLDNPNKKAEETIKYEQQKRIIFI 70

Qy 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYSPRDKQIRFMEEN-SLDFSATLTIEL 122

Db 71 KNERNLGRGASRNKAVXIARGKYIALLDADDIALPKRLKQFKYMNENRDIILFSWYF 130

Qy 123 IDOKNLVYKQESNKVYLTNDIRKMLNRSILAHPTWCVKKKVFDKLMGYRDLVPVEDY 182

Db 131 IDENGNIL-KFKPKPK-YKFEIKKYFFKEHLTVAPSMVMSKILKKLYDKELRSQDY 188

Qy 183 DFAIRGALADFKIGLNLKVLQYRL--NENGISQTNKFKQYIYSAILODFYKESY---- 236

Db 189 DFWIRCIANDYKFDIEEFLLKRYIPNRDNYLSRIKKQKLSYYT-LKTHWNKKHFCNN 247

Qy 237 IDITKITYFOEYVKKRYTQOELSXYFELKSTPSITIRKL 277

Db 248 VYFWKV--FFYSLVV-----YLFIVLTPTFILKIL 275

RESULT 2

H97167

glycosyltransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: H97167
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
. J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bac
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: H97167
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-333 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80131.1; PID:gl5025167; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2173

Query Match 23.1%; Score 367; DB 2; Length 333;
Best Local Similarity 33.3%; Pred. No. 1.6e-16;
Matches 93; Conservative 62; Mismatches 92; Indels 32; Gaps 10;

[illegible]

QY 238 DITKITNYFQEYVIKKRY-----TQQELSKYFE 265
 | | : : : :
 : : : : : :
 Db 230 KIKLIDVFKREFKDKFYIWMGASNGGKITKEVLDEFFE 268

```

RESULT      3
A97168
glycosyltransferase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97168
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bact
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <KUR>
A:Cross-references: GB:AF001437; PIDN:AAK80132.1; PID:g15025168; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2174

```

Query Match 21.9%; Score 348.5; DB 2; Length 336;
Best Local Similarity 36.6%; Pred. No. 2.5e-15;
Matches 90; Conservative 46; Mismatches 95; Indels 15; Gaps 8;

	Qy	4	SIIMSVYNEPLNYVRDVSVESTLNOTLTDFETIIVIDNPISRGDLKQFLFYSVVDNRKIL	63
			: : : : : : : : :	
	Dd	8	SIVMPVYNSE-RYLAAEATESISLDQTYNDFETIIV-DDGSTDSEYNI ISSYANKONRI -IV	64
			: : : : : : : : :	
	Qy	64	LNEENIGLASSLNRAVKITSGKEYIFRMDADDISYPSRFEDKOIRFMEEN-SLDFSATLIEL	122
			: : : : : : : : :	

C; Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C; Accession: J97367
R; Nollung, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A; Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Citrobacter*
A; Reference number: A96900; MUID: 21359325; PMID: 21359325

A:Accession: H97167
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <KUR>
A:Cross-references: GB:AE001437; PID:AAK80131.1; PID:g15025167; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2173

Query Match	23.1%	Score 367;	DB 2;	Length 333;
Best Local Similarity	33.3%	Pred. NO. 1.6e-16;		
Matches 93;	Conservatives 62;	Mismatches 92;	Indels 32;	Gaps 10;

Qy	4	SIIMSVYNEPNLYVRDVSVESTILNLTOTLDFEPIIIVDNPISRGDLKQFLTETYSVVDNRKIL	63
Db	5	SVWVPYVNSE-KYLKTESITLNQSYSDLEPIIINDGSTONSEK-IIKEYAKLDRINV-	61
Qy	64	LNEENIGLASSLNKAVKISKEYIFRWADDISYPSFEDKQIREMBEN-SIDFSATLIEL	122
Db	62	ISRENKGIVSLNRAIRAKGEYIARMADDISAPKRIEKQISFLKSHRDIILGTQVKV	121
Qy	123	IDQRGNLYVKQRESNKIYLTNDI---RKMLNLR-SILAHPTWCVKKKVDFKLMGYRDLV	177
Db	122	VGNISNDI-KEKNENKLNIEFDYDDNREKILNYWCLAHPVSMFRKDILRELKCYNDF-	179
Qy	178	PVEDYDFAIRGALADFKIIGLNLKVLLOYLRNLNENGISOTNFKQVIYSAILQDFYKESYI	237
Db	180	KSEDLJWLRAIESGFKYIKLKEELIIVFRHEESKTRVDN-----ONYEGLKXGI	229

```
Qy 238 DITKITNYFOEYVIKKRY-----TOQELSKYFE 265
      |   |   |   |   |   |   |   |   |   |
Db 230 KILIDVPKREFKDKFYIVMGASNGGKIKTEVLDEFFE 268
      |   |   |   |   |   |   |   |   |   |
```

RESULT 3
A97168
glycosyltransferase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97168
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <KUR>
A:Cross-references: GB:AE001437; PTDN:AAK80132.1; PID:gl5025168; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2174

Query Match	21.9%	Score 348.5;	DB 2;	Length 336;
Best Local Similarity	36.6%	Pred. No. 2.5e-15;		
Matches 90;	Conservative 46;	Mismatches 95;	Indels 15;	Gaps 8;

[illegible]

Db	65	ISREHRLGVLDSLNEGNIARGKVIARMADDISINNRIEKQFFLELNKQDVLGTRIEA	124
Qy	123	---TDQGNLVYQKQESNKIYLTNDIRKMLNRSILAHPTWCVKKKVFQKLMGYR-DLVP	178
Db	125	FGDIDEKQKIYNSAFSIK-FDSQTEQVPLTSCAIPHFSVMPFKKDSIVKLGRYKEYDT	183
Qy	179	VEDYDFAIRGALADFKLGLLNKVLLOYRLNENGISQTNFKQIYVSAILQDFYKEKSYID	238
Db	184	AEDYDLWLRAIRNGYKIVRWDECLIKYRVHNSKTAVERNPKQWE-----YTMKAKID	237
Qy	239	ITKTN	244
Db	238	YINDT	243

RESULT 4
B971168
glycosyltransferase [Imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B971168
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gib-
.: Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing
A:Reference number: A96900; MUID:21359325
A:Accession: B971168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK60133.1; PID:g15025169; GSPDB:GN000
A:Experimental source: Clostridium acetobutylicum AFCC824
C:Genetics:
A:Gene: CAC2175

Query Match 19.9%; Score 317; DB 2; Length 333;
Best Local Similarity 32.8%; Pred. No. 2.6e-13;
Matches 83; Conservative 53; Mismatches 91; Indels 26; Gaps 10;

Qy	4	SIIMSVYNEPLNYVRDVS ESILNQTLTDFEII VIDNPSRDLKQFLT EYSVDNRKIL	63
Db	3	SVIMPVYNCE-KYLE ESIESILKQTYRD FEFIVNDGSDNKSI- DIINKYANDNRIVV	60

QY		64	LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPREDKQIRMEEN-SLDFSATLIEL	122
	:	1-1:	111: : 11 : 111111: 11:11 : : :	:
Dd		61	SRDNNMGVYSLNEGIDRAKGSYVARNDADDTALPERFERQIEYLNKNKDVDILACKVEA	120

QY 123 IDQGNLVVQKRESNKIYLTND-----IRKMLLNRSILAHPTWCVKKKVFDKLMGYR-D 175
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 121 F----GDVSREQKLEHRYNVVDLNNSESIESFLENCTIAHPSVMMKMMSVLKALGGYNLN 177

QY 176 LVPVEDYDFAIRGALADFKICLLNKKVLLQYRL-NENGISQT- ---NKFKQYIYSAILQDF 230
 178 YKRTEDYNLWLRATAGYKIAMLEEKIMKIRLHNSDKTHRDAGFSSIRDIIOSSL--EY 235

QY 231 YKEK-----SYI 237
||| ||:
db 236 VKELIKLKDFSYV 248

RESULT 5
B84114
exopolysaccharide biosynthesis BH3714 [imported] - Bacillus halodurans (str
B84114)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: B84114
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; F
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus ha
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: B84114
A:Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-271 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07433.1; GSPDB:GN000000000
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3714

Query Match      18.9%; Score 300.5; DB 2; Length 271;
Best Local Similarity 36.6%; Pred. No. 2.3e-12;
Matches 82; Conservative 41; Mismatches 92; Indels 9; Gaps 6;

Qy      4  SIIMSVNEPLNYVRDSVESILNQTLDTFEPIIVDNPSCRGLKQFLTETYSVVDNRKIL 63
      ::::: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db      6  TVLMSVYNDK- NYLSESIENLTNFTENFEFLIINDASTDGS- GELLEYSKKDKIRLI 63

Qy      64  LNEENIGLASSLANKAVKISGEYIFRMDADDISYPSRFDKQIREMEENS- LDFSATLIEL 122
      |: | ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db      64  HNKNRGLSVSLAGSVSLAKAPWITARMADDVSPKRLAVQMHDVKAHSELDILIGSVVD 123

Qy      123  IDQAGNLVYQRESNKLYLTINDIRKMLLNRSILAHPTWCVKKKVFDKLMGY- RDLVPVED 181
      || || ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db      124  IDDKGN- - - -ELEIRKVPYTHKEANLIWTCPTHPTVLEKKDSIIKAGSVDRNLRRQD 179

Qy      182  YDFAIRGALADFKTLGLNKKVLLQYLRNLNENGISQPNKFKQYIYA 225
      || || ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db      180  YDLWFRCLAKLKFENIDKPLLYRSTDD- YKKKNFKVVOOQA 222

```

RESULT 6
T44330
glycosyl transferase homolog [imported] - *Vibrio cholerae*
C/Species: *Vibrio cholerae*
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C/Accession: T44330
R/Yamasaaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Nair, G.B.; Takeda, Y.
Gene 237, 321-332, 1999
A/Title: the genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are close
A/Reference number: 222749; MUID:99453293
A/Accession: T44330
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-340 <YAM>
A/Cross-references: EMBL:AB012957; NID:q4115688; PIDN:BAA33634.1; PID:q3721684
A/Experimental source: strain O22
C/Genetics:
A/Note: wblc
C/Superfamily: *Neisseria meningitidis* glycosyl transferase A

Query Match	17.7%	Score 282;	DB 2;	Length 340;
Best Local Similarity	36.9%;	Pred. No. 4.6e-11;		
Matches 82;	Conservative 35;	Mismatches 77;	Indels 28;	Gaps 10;
Qy	4	SIIMSVNEPLNYVRDSVESILNGLTDFEIIIVDNPISRGDLKQFLTEYSVDNRKIL 63		
Db	9	SVIMSYNGE-KYLAQAIESILNGLTFSDFEIIVDDGSTDSL-SIIQAYMDKDRI-VL 65		
Qy	64	LNERNIGLASSLNKAVISKGEYITFRMDADDISVPSRFDKQIRMEENSLDFSATLIELI 123		
Db	66	ISRVNKGPLYSLENAISVSKANYIARDADDISLPERLETQLAYMENNP-DIGVC----- 119		
Qy	124	DQKNLNVKQRES---NKIYLTNDIRKMLNRSILA-----HPTWCVKKKVPDKLIMGYRDL 176		
Db	120	---GTLAYLFRETSPKNNKMCHPDHDSLIIRLLFVCFIHPVVMIRKSVLDQL-----DY 172		
Qy	177	VPVEDY-----DFATRGALAD-FKIGLLNKVLLQYRLNENGIS 213		
Db	173	VYNENFRNSQDYELWSRIAEXTREYTIQKPLFFYRDTPDGT 214		

RESULT 7
G95948

probable glycosyltransferase protein SMB211189 [imported] - *Sinorhizobium meliloti* (str C; Species: *Sinorhizobium meliloti*)
C; Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C; Accession: G95948
R; C; Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her-
proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A; Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing e-
A; Reference number: A95842; MUID:2136508; PMID:11481431
A; Accession: G95948
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-367 <R>
A; Cross-references: GB:AL591985; PIDN:CAC49255.1; PID:g15140741; GSPDB:GN00167
A; Experimental source: strain 1021, megaplasmid pSymb
R; Gallibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl-
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 283, 668-672, 2001
A; Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Letlau-
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A; Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A; Reference number: A96039; MUID:21368234; PMID:11474104
A; Contents: annotation
C; Genetics:
A; Gene: SMB211189
A; Genome: plasmid

Query Match	17.08;	Score	271;	DB 2;	Length	367;			
Best Local Similarity	33.88;	Pred. No.	2.6e-10;						
Matches	74;	Conservative	52;	Mismatches	79;	Indels	14;	Gaps	11;
Qy	4	SIIMSVYN-nPLNVVRDVSVESILNQTLDTEFEIVIDNPNSRGDLKQFLTEYVVDNRKI	62						
Db	7	SIVLPVNAEP--VIAAAIESVLQDYELE-VTAIDGSDTRSDILERYKSDSRVSI	63						
Qy	63	LLNEENTGLASSLNKAVKIKSGEYIFRMDADDISYPSRFDKQIR-FMEENSILDSATLIE	121						
Db	64	-ISRENGLIATLNEGALAKGELIARMADDIAIYPSRLSRQVALFSAEPRLALSGTGID	122						
Qy	122	LIDQKGNLVKQRESNKIYLTNDIRKMLLRSILAHPTWCVKKKVF-DKLMG--RDLPVP	179						
Db	123	ML--IGNRIIRGK--PNPIYRPGSLRILSMFTTIPMHSTVYVYNNRNVIPDEMLRDPNVHA	179						
Qy	180	EDYDPAIRGALAD-FKIGLLNKVLQYRLNENGISQTNK	217						
Db	180	EDFDLFR--IADRFPMHIMIDEALVAYRIHEDSVTSKHK	216						

RESULT 8

glycosyl transferase homolog H1578 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C/Accession: H64130
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae rd.
A/Reference number: A64000; MUID:95350630
A/Accession: H64130
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-323 <TIGR>
A/Cross-references: GB:U32832; GB:LA2023; NID:q1574421; PIDN:AAC23227.1; PID:q1574422
C/Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match	16.6%	Score 264;	DB 1;	Length 323;
Best Local Similarity	29.0%	Pred. No. 6.2e-10;		
Matches 93;	Conservative 62;	Mismatches 114;	Indels 53;	Gaps 13;

E97757
hypothetical protein RC0461 [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97757
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: E97757
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-604 <KUR>
A:Cross-references: GB:AB006914; PIDN:AAL02999.1; PID:g15619533; GSPDB:GN00173
C:Genetics:
A:Gene: RC0461

Query Match 13.8%; Score 220; DB 2; Length 604;
Best Local Similarity 25.8%; Pred. No. 9e-07;
Matches 77; Conservative 65; Mismatches 118; Indels 38; Gaps 10;

QY 4 SIIMSVNEPLNVYRDSVESILNQTLDPEFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
DB 296 SVIIPVYNR-INWAEIAKSVLIQTHKPE-ILIIDGSDTDISE-LTAICKKDKRIK-Y 351
QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEENSLDPSATLIELI 123
DB 352 FHKKNEGPAARNLIGKNAIGXYIAFLSDDLFYKDKIEIQLKFMEEENNFISSHYSYHKI 411
QY 124 DQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTWCYKKVFDKLMGYRDLVPVED-- 181
DB 412 NEKGKVIESHVSG--LFSGNVPQVICTPIAMPTVMGTLTLFQENLFPENIRSGEDCC 468
QY 182 --YFAIRCALA-----DFKIGLKNKVLLOYRLNENGISQTNK 217
DB 469 LMSIASKNSIGGIDKELSKVRISGNTFTMDPNKYSVGLIN--ITSVYLNDAIYLSKFSP 526
QY 218 FK-QVIYSAILQDFYKEKSYIDITK--ITNYPQEVYKRVYQQBELSKYFELKSTPSI 272
DB 527 FTINLLAAVTOQLLENKNEDYKKSNI5FFKNVYIQIRTYCFVTKILILLTITSI 584

RESULT 15
AH2026
hypothetical protein all1766 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AH2026
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2026
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW73465.1; PID:g17130856; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1766

Query Match 13.7%; Score 218.5; DB 2; Length 333;
Best Local Similarity 23.9%; Pred. No. 5.3e-07;
Matches 75; Conservative 70; Mismatches 120; Indels 49; Gaps 13;

QY 4 SIIMSVNEPLNVYRDSVESILNQTLDPEFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
DB 12 SVIIPAYNCE-KTIKKTIDSVLDQSFTDFELIVINDGSQDATLD---IVSQIEDSRIKI- 66

QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN-SLDPSATLIEL 122
DB 296 SVIIPVYNR-INWAEIAKSVLIQTHKPE-ILIIDGSDTDISE-LTAICKKDKRIK-Y 351
QY 123 IDOKNLVYKQRESNKIYLTNDIRKMLNRSIL---AHTWCYKKVFDKLMGY-RDLVP 178
DB 127 IDEDNFLISGR---RVTLNGDVYKKLFINNFLENGSNPLIC--KEALIALGGFDESLSKA 181
QY 179 VEDYDFAIRGALADPKIGLLNK-----VLLQYRLNENGISQTNKQYIYSAILODF 230
DB 182 AQDWMWLR-----LANKYSFVAVPYVOILYRVSSNSLSSNLVROEKACKQVLEKA 232
QY 231 YKEKS-----YIDITITNYPQEVYKRVYQQBELSKYFELKSTPSITIRK 276
DB 233 YQARPAIGNHILHLSTIANLYKYLACKALQKPYSRSGKGLAAVKFIWQYFLYDSS---RMRR 289
QY 277 LYICILYLYFKSPLV 290
DB 290 INSTKLKLLFKALII 303

Search completed: July 31, 2002, 18:56:46
Job time: 342 sec

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	210.5	13.2	348	1	US-08-312-387B-3	Sequence 3, Appli	
2	210.5	13.2	348	1	US-08-312-387B-11	Sequence 11, Appli	
3	210.5	13.2	348	1	US-08-683-426-3	Sequence 3, Appli	
4	210.5	13.2	348	1	US-08-683-426-11	Sequence 11, Appli	
5	210.5	13.2	348	1	US-08-683-458-3	Sequence 3, Appli	
6	210.5	13.2	348	1	US-08-683-458-11	Sequence 11, Appli	
7	210.5	13.2	348	1	US-08-878-360-3	Sequence 3, Appli	
8	210.5	13.2	348	2	US-08-878-360-11	Sequence 11, Appli	
9	210.5	13.2	348	3	US-08-478-140B-3	Sequence 3, Appli	
10	210.5	13.2	348	3	US-08-478-140B-8	Sequence 8, Appli	
11	210.5	13.2	348	4	US-09-333-412-3	Sequence 3, Appli	
12	210.5	13.2	348	4	US-09-333-412-11	Sequence 11, Appli	
13	176.5	11.1	337	1	US-08-312-387B-5	Sequence 5, Appli	
14	176.5	11.1	337	1	US-08-312-387B-12	Sequence 12, Appli	
15	176.5	11.1	337	1	US-08-683-426-5	Sequence 5, Appli	
16	176.5	11.1	337	1	US-08-683-426-12	Sequence 12, Appli	
17	176.5	11.1	337	1	US-08-683-458-5	Sequence 5, Appli	
18	176.5	11.1	337	1	US-08-683-458-12	Sequence 12, Appli	
19	176.5	11.1	337	2	US-08-878-360-5	Sequence 5, Appli	
20	176.5	11.1	337	2	US-08-878-360-12	Sequence 12, Appli	
21	176.5	11.1	337	3	US-08-478-140B-5	Sequence 5, Appli	
22	176.5	11.1	337	3	US-09-333-412-5	Sequence 5, Appli	
23	176.5	11.1	337	4	US-09-333-412-12	Sequence 12, Appli	
24	161	10.1	324	1	US-08-597-236-10	Sequence 10, Appli	
25	161	10.1	324	1	US-08-746-682A-10	Sequence 10, Appli	
26	160	10.1	281	4	US-08-961-083-196	Sequence 196, App	
27	129.5	8.1	303	4	US-08-961-083-202	Sequence 202, App	

; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-140B-3

Query Match 13.2%; Score 210.5; DB 3; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY 4 STIMSYNEPLNYVRDSEVSIINQTLTDFEFLIVDNPGRDLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYFAQSLAAVYVQWNRNLDILIVDDGSDGTL-AIAKDFQKRDRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQOQNSGLTSPSINIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT----NDIRKMLNRSILAHPTWCVKKVFOKL 170
Db 124 GAWLEVLSEKDGRLARHHKHKIWKKPTRHEDIAAAPPFGNPIHNTMTIMRRSVIDGG 183
QY 171 MGY----RLVPVEDYDF-----AIRGALADFKIGLKNKVLQYRLNENGISQTNKFKQIYI 223
Db 184 LRYDTERDW--AEDYQFWDVSKLGRLAYP-----EALVKYRLHANQVSSKHSVRQH-- 234
QY 224 SAILQDFYKEKSYIDITKITNYFOEYVKKRYT---QOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAQGIQK-----TARNDFLOSNGFKTRFDSLEYRQTAKAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFYQCFK 295

RESULT 10
US-08-478-140B-8
; Sequence 8, Application US/08478140B
; Patent No. 6127153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; APPLICANT: BUCZALA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ANTI-SENSE: NO
US-08-478-140B-8

Query Match 13.2%; Score 210.5; DB 3; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY 4 STIMSYNEPLNYVRDSEVSIINQTLTDFEFLIVDNPGRDLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYFAQSLAAVYVQWNRNLDILIVDDGSDGTL-AIAKDFQKRDRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQOQNSGLTSPSINIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT----NDIRKMLNRSILAHPTWCVKKVFOKL 170
Db 124 GAWLEVLSEKDGRLARHHKHKIWKKPTRHEDIAAAPPFGNPIHNTMTIMRRSVIDGG 183
QY 171 MGY----RLVPVEDYDF-----AIRGALADFKIGLKNKVLQYRLNENGISQTNKFKQIYI 223
Db 184 LRYDTERDW--AEDYQFWDVSKLGRLAYP-----EALVKYRLHANQVSSKHSVRQH-- 234
QY 224 SAILQDFYKEKSYIDITKITNYFOEYVKKRYT---QOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAQGIQK-----TARNDFLOSNGFKTRFDSLEYRQTAKAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFYQCFK 295

RESULT 11
US-09-333-412-3
; Sequence 3, Application US/09333412
; Patent No. 6342382
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,412
; FILING DATE: 15-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: July 7, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEX: 201 343-1684
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-333-412-3

Query Match 13.2%; Score 210.5; DB 4; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;
QY 4 SIIMSVNEPLNYVRDSVESILNOTLTDFFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYPAQSLAAVWNTWRNLDLIVDDGSTDGTL-AIAKDFKRDRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQONSLIPSLNIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT----NDIRKMLNRSILAHPTWCVKKKVDFKL 170
Db 124 GAWLEVLSEKDGRLARHHKHGKIWKKPTRHEDIAAFFPGNPIHNTMTMRRSVIDGG 183
QY 171 MGY---RDLVPVEDYDF-----AIRGALADPKIGLLNKVLQYRLNENGISOTNKFQYIY 223
Db 184 LRYDTERDW-AEDYQFWDVSKLGRLAYP-----EALVKYRLHANQVSKKHSVRQH-- 234
QY 224 SAILQDPYKESYIDITKITNYFOEYVYKRYT----QOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAQGIQK-----TARNDFLQSMGFKTRFDSLEYRQTAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFYQCCK 295

RESULT 12
US-09-333-412-11
; Sequence 11, Application US/09333412
; Patent No. 6342382
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/333.412
; FILING DATE: 15-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER: 08/312,387
; FILING DATE: July 7, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEX: 201 343-1684
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-333-412-11
Query Match 13.2%; Score 210.5; DB 4; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;
QY 4 SIIMSVNEPLNYVRDSVESILNOTLTDFFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYPAQSLAAVWNTWRNLDLIVDDGSTDGTL-AIAKDFKRDRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQONSLIPSLNIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT----NDIRKMLNRSILAHPTWCVKKKVDFKL 170
Db 124 GAWLEVLSEKDGRLARHHKHGKIWKKPTRHEDIAAFFPGNPIHNTMTMRRSVIDGG 183
QY 171 MGY---RDLVPVEDYDF-----AIRGALADPKIGLLNKVLQYRLNENGISOTNKFQYIY 223
Db 184 LRYDTERDW-AEDYQFWDVSKLGRLAYP-----EALVKYRLHANQVSKKHSVRQH-- 234
QY 224 SAILQDPYKESYIDITKITNYFOEYVYKRYT----QOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAQGIQK-----TARNDFLQSMGFKTRFDSLEYRQTAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFYQCCK 295
RESULT 13
US-08-312-387B-5
; Sequence 5, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/312.387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 18:50:19 ; Search time 30.44 seconds
(without alignments)
1142.119 Million cell updates/sec

Title: US-09-900-038a-1

Perfect score: 1590

Sequence: 1 MNYSLMSVYNEPLNVRDS.....LINDINILVLKFGGEKQSD 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272	17.1	268	21 AAY54093	Enzyme EPSH involv
2	272	17.1	268	21 AAY43795	Amino acid sequenc
3	240.5	15.1	274	22 AAG90151	C glutamicum prote
4	234	14.7	301	21 AAY97206	Campylobacter jeju
5	223	14.0	209	21 AAY68985	Amino acid sequenc
6	216	13.6	706	22 AAU33454	Enterococcus faeca
7	214	13.6	715	22 AAU34918	Enterococcus faeca
8	216	13.5	269	21 AAY68980	Cps9C protein whic
9	213	13.4	260	20 AAW88312	Sugar transferase
10	210.5	13.2	348	17 AAK91311	N. gonorrhoeae gly
11	210.5	13.2	348	18 AAW06576	Neisseria polyglyc

12	210	13.2	322	21 AAY68974	CpsII protein whic
13	200	12.6	270	21 AAY97203	Campylobacter jeju
14	198	12.5	298	22 AAB96313	Putative glycosylt
15	193.5	12.2	327	21 AAY54095	Enzyme EPSJ involv
16	193.5	12.2	327	21 AAY43797	Amino acid sequenc
17	187	11.8	332	21 AAY68962	Cps2J protein whic
18	186.5	11.7	345	22 AAG90140	C glutamicum prote
19	182.5	11.5	329	22 AAB47427	ESPM. Lactococcus
20	180.5	11.4	302	21 AAY97213	Campylobacter jeju
21	176.5	11.1	337	17 AAY91314	N. gonorrhoeae gly
22	176.5	11.1	337	18 AAW06579	Lipo-oligosacchari
23	168	10.6	277	19 AAW80728	P. pneumoniae gly
24	168	10.6	702	21 AAY96212	P. multocida hyalu
25	165.5	10.4	972	20 AAY06212	Pasteurella multoc
26	165.5	10.4	972	21 AAY43099	P. multocida hyalu
27	165	10.4	334	21 AAY68963	Cps2K protein whic
28	165	10.4	336	21 AAY54098	Enzyme EPSM involv
29	165	10.4	336	21 AAY43800	Amino acid sequenc
30	165	10.4	965	21 AAY96213	P. multocida chond
31	164	10.3	278	21 AAY68976	CpsIK protein whic
32	161	10.1	324	18 AAW14078	S. thermophilus exo
33	161	10.1	324	18 AAW22177	S. thermophilus exo
34	161	10.1	328	21 AAY81720	Streptococcus pneu
35	160	10.1	281	19 AAW61236	Streptococcus pneu
36	159.5	10.0	332	22 AAB47426	EPSN. Lactococcus
37	158.5	10.0	330	21 AAY54080	Enzyme EPS6 involv
38	158.5	10.0	330	21 AAY43782	Amino acid sequenc
39	157.5	9.9	322	21 AAY68975	CpsIJ protein whic
40	156.5	9.8	322	21 AAY54071	Enzyme EPS7 whic
41	156.5	9.8	322	21 AAY43773	Amino acid sequenc
42	150	9.4	973	22 ABG02248	Novel human diagno
43	150	9.4	973	22 ABG13176	Novel human diagno
44	150	9.4	973	22 ABG25145	Novel human diagno
45	147	9.2	316	21 AAY54072	Enzyme EPS8 whic

ALIGNMENTS

```

RESULT 1
AAY54093
ID AAY54093 standard; Protein; 268 AA.
XX
AC AAY54093;
XX
DT 27-MAR-2000 (first entry)
XX
DE Enzyme EPSH involved in exopolysaccharide biosynthesis.
XX
KW Exopolysaccharide; EPS; ESP enzyme; EPSA; EPSB; EPSD; EPSE; EPSF;
KW EPSG; EPSH; EPSJ; EPSK; Lactobacillus delbrueckii bulgaricus;
KW activated D-galactose pyranose; saccharide; beta-glycosyltransferase;
KW transcription attenuator; glucosyl-phospho-transferase;
KW alpha-glycosyltransferase; glucosyltransferase; EPSJ; EPSM;
KW alpha-glycosyltransferase; EPS polymerase; glycosyltransferase; EPSN;
KW phosphofuranose; transporter; food; fermented milk product; yoghurt;
KW cheese; flavour stability; organoleptic property.
XX
OS Lactobacillus delbrueckii bulgaricus.
XX
PN WO9962316-A2.
XX
PD 09-DEC-1999.
XX
PF 22-APR-1999; 99WO-EP02841.
XX
PR 22-APR-1998; 98EP-0201310.
PR 22-APR-1998; 98EP-0201311.
PR 22-APR-1998; 98EP-0201312.
XX
PA (NEST ) SOC PROD NESTLE SA.
XX
PI Stingele F, Germond JE, Lamothe G;

```

XX WPI; 2000-097267/08.
 DR N-PSDB; AAZ54260.
 XX
 PT New recombinant enzymes for synthesis of exopolysaccharides,
 PT particularly in lactic acid bacteria, for improving properties of
 PT fermented milk products -
 XX
 PS Claim 13; Page 154-155; 162pp; French.
 XX
 CC AAY54086-99 represent enzymes involved in the biosynthesis of
 CC exopolysaccharides (EPS). These enzymes are designated EPSA-EPSN, and
 CC are encoded by open reading frames epsA-epsN. The enzymes are isolated
 CC from Lactobacillus delbrueckii bulgaricus. The proteins are used
 CC in a method for the synthesis of EPS, which includes at least one step
 CC of forming a bond (alpha or beta-isomer) between C-1 (carrying the
 CC reducing aldehyde function, of an activated D-galactose pyranose), and
 CC a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis
 CC of EPS occurs with, in each step, addition of a new sugar unit, through
 CC its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar
 CC unit, present at the end of a chain of sugar residues bonded to the
 CC primer. EPSA is attenuator of transcription which influences
 CC regulation; EPSB determines the length of the saccharide chain; EPSK is
 CC involved in the control of the molecular weight and/or the length of the
 CC polysaccharide; EPSD, EPSL and EPSM are involved in synthesis of EPS;
 CC EPSL is a galactosyl- or glucosyl-phospho-transferase which catalyses the
 CC transfer of the first saccharide on the primer; EPSF and EPSG are
 CC alpha-glucosyltransferases; EPSH and EPSHJ are
 CC beta-glucosyltransferases; EPSI is a glucosyltransferase; EPSK is
 CC responsible for the polymerisation of the repetitive units; and EPSN is
 CC responsible for export of the EPS. The EPS enzyme are used to improve
 CC properties of foods, particularly fermented milk products such as yoghurt
 CC and cheese, e.g. their organoleptic properties and flavour stability.
 XX
 SQ Sequence 268 AA;

Query Match 17.1%; Score 272; DB 21; Length 268;
 Best Local Similarity 29.7%; Pred. No. 5.3e-17;
 Matches 82; Conservative 56; Mismatches 102; Indels 36; Gaps 11;

QY 4 STIMSVYNEPLNY--VRDSVESILNQTLDFFFIIVID-NPSRGDLKQFLTEYSVVDNRI 60
 Db 5 svmsiyn-cnkylakelsieslidqtsdweifiyndgsddgktsyellklgkrdsr 63
 QY 61 KILLNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN-SLDFSATL 119
 Db 64 qliidcpnnhglayaknemikvargdyitaqdddvshpmrlarevdfinrhpayaavgv 123
 QY 120 IELIDQKGNLVY---KQESNKIYLTNDIRKMLLNRSILAHPTWCVKKKVFKLMGYR-- 174
 Db 124 akvfdnsgtgwhgyikeptkntflwn-----spflhpsvmfrhevldqvgngy 174
 QY 175 -DLVPVEDYDFAIRGALADFKIGLLNKLVLQYRLNENGISQTNKFKQYIYSAILQDFYKE 233
 Db 175 kdtmraedydlfrliyakgkyniqedleyeriennpnkkyrpsmdriqekavr--ykg 232
 QY 234 -KSYIDIRKITNYFOE-----YVIKK-RY 255
 Db 233 fksfgmltkgipylkplliglipkqifirknry 268

RESULT 2
 ID AAY43795
 XX AAY43795 standard; Protein; 268 AA.
 AC AAY43795;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of epsH of L. delbrueckii bulgaricus Lfi5.
 XX
 KW eps operon; Lactobacillus delbrueckii bulgaricus Lfi5; enzyme; epsA;

KW epsB; epsC; epsD; epsE; epsF; epsG; epsH; epsI; epsJ; epsK; epsL; epsM;
 KW exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;
 KW probiotic; foodstuff; organoleptic quality; flavour;
 XX lactic acid bacteria; acidified milk product; yoghurt; cheese.
 OS Lactobacillus delbrueckii bulgaricus.
 XX
 PN WO9954475-A2.
 XX
 PD 28-OCT-1999.
 XX
 PF 22-APR-1999; 99WO-EP03011.
 XX
 PR 22-APR-1998; 98EP-0201310.
 PR 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 PI Stingele F, Germond JE, Lamothe G;
 DR WPI; 2000-013255/01.
 DR N-PSDB; AAZ30357; AAY43796, AAY43797, AAY43798, AAY43799, AAY43800,
 DR AAY43801, AAY43842.
 XX
 PT New recombinant enzymes for biosynthesis of exopolysaccharides having
 PT e.g. antitumor or probiotic properties or useful in fermented milk
 PT products -
 XX
 PS Claim 13; Page 155-156; 163pp; French.
 XX
 CC AAY43788-89, AAY43791-Y437801 and AAY43842 represent the enzymes encoded
 CC by the eps operon of Lactobacillus delbrueckii bulgaricus Lfi5. The
 CC operon contains 14 open reading frames, and encodes enzymes (epsA,
 CC epsB, epsC, epsD, epsE, epsF, epsG, epsH, epsI and epsJ, epsK, epsL,
 CC and epsM) that are involved in the biosynthesis of exopolysaccharides
 CC (EPS). The enzymes catalyse the formation of specific intersugar bonds.
 CC The enzymes catalyse a process that includes at least one step of
 CC forming a bond (in alpha or beta anomeric form) between C1, carrying
 CC the reducing aldehyde group of an activated D-Galp (galactose in
 CC pyranose form), and a phosphate group on a lipophilic or proteinaceous
 CC primer. The enzymes are used to produce EPS that have antitumor or
 CC probiotic properties or are used in foodstuffs to improve organoleptic
 CC qualities and flavour. When expressed by lactic acid bacteria, EPS
 CC impart a free-flowing character and/or a smooth, creamy texture to
 CC acidified milk products (yoghurt or cheese).
 XX
 SQ Sequence 268 AA;

Query Match 17.1%; Score 272; DB 21; Length 268;
 Best Local Similarity 29.7%; Pred. No. 5.3e-17;
 Matches 82; Conservative 56; Mismatches 102; Indels 36; Gaps 11;

QY 4 STIMSVYNEPLNY--VRDSVESILNQTLDFFFIIVID-NPSRGDLKQFLTEYSVVDNRI 60
 Db 5 svmsiyn-cnkylakelsieslidqtsdweifiyndgsddgktsyellklgkrdsr 63
 QY 61 KILLNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN-SLDFSATL 119
 Db 64 qliidcpnnhglayaknemikvargdyitaqdddvshpmrlarevdfinrhpayaavgv 123
 QY 120 IELIDQKGNLVY---KQESNKIYLTNDIRKMLLNRSILAHPTWCVKKKVFKLMGYR-- 174
 Db 124 akvfdnsgtgwhgyikeptkntflwn-----spflhpsvmfrhevldqvgngy 174
 QY 175 -DLVPVEDYDFAIRGALADFKIGLLNKLVLQYRLNENGISQTNKFKQYIYSAILQDFYKE 233
 Db 175 kdtmraedydlfrliyakgkyniqedleyeriennpnkkyrpsmdriqekavr--ykg 232
 QY 234 -KSYIDIRKITNYFOE-----YVIKK-RY 255
 Db 233 fksfgmltkgipylkplliglipkqifirknry 268

RESULT 3
 ID AAG90151 standard; Protein; 274 AA.
 XX
 AC AAG90151;
 XX
 XX 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 3905.
 DE
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 KW
 XX Corynebacterium glutamicum.
 OS
 XX EP1108790-A2.
 PN
 XX 20-JUN-2001.
 PD
 XX 18-DEC-2000; 2000EP-0127688.
 XX
 PF 16-DEC-1999; 99JP-0377484.
 XX
 PR 07-APR-2000; 2000JP-0159162.
 PR
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 PI
 XX WPI: 2001-376931/40.
 DR
 DR N-PSDB; AAH65370.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 XX Claim 17; SEQ ID NO: 3905; 246pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 274 AA;

Query Match 15.1%; Score 240.5; DB 22; Length 274;
 Best Local Similarity 26.2%; Pred. No. 4.5e-14;
 Matches 71; Conservative 72; Mismatches 101; Indels 27; Gaps 10;

QY 10 YNEPLNVRDVSILNQTLTDFEIIIVIDNPSRGDLKQFLTEYSVDNRKILLNENI 69
 Db 17 yde---ycsqsksvceqyenqgivilvdgaplkdvpqwkheh-----erikiveqirg 69
 QY 70 GLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEENS-LDFSATLIELIDOKGN 128
 Db 70 lvtstlnngkiasdgqllardldsdlaapsrlskqeeflnrhpyilcvactkthinehgk 129
 QY 129 LVYQKRESNKIYITNDIRKMLLSRLSLAHTPTWCVKKKVFDKLMGYR-DLVPVEDYDFAIR 187

Db 130 if---ggsadlptsqdirqillvknpihasvmyrkqvveqigysiemtrsqdyelflr 186
 QY 188 ----GALADFKIGLLNKVLQYRLNENGIS-OTNKFQOYIYSAILODFYKEKSYIDITKI 242
 Db 187 lskiga-----igyldeslsyringqhsrktspfkkytw-iilkrmelasfikrspv 240
 QY 243 TNYFQEVYIKKRYTQOELSRYFELKSTPSIT 273
 Db 241 rqiflnfi---wygaqvtrylgirkagfmt 267

RESULT 4
 AAY97206
 ID AAY97206 standard; Protein; 301 AA.
 XX
 AC AAY97206;
 XX
 XX 22-DEC-2000 (first entry)
 DT
 DE Campylobacter jejuni OH4384 Beta-1,3-galactosyltransferase.
 XX
 XX Biosynthetic locus; biosynthesis; lipid A biosynthesis;
 KW acetyltransferase; glycosyltransferase; Beta-1,4-GalNAc transferase;
 KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
 KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
 KW immunity; immunogen; ganglioside.
 XX
 OS Campylobacter jejuni OH4384.
 XX
 XX WO2000046379-A1.
 PN
 XX 10-AUG-2000.
 PD
 XX 01-FEB-2000; 2000WO-CA00086.
 PF
 XX 01-FEB-1999; 99US-0118213.
 PR
 PR 31-JAN-2000; 2000US-0495406.
 XX
 XX (CANA) NAT RES COUNCIL CANADA.
 PA
 XX Gilbert M, Wakarchuk WW;
 PI
 XX WPI: 2000-524418/47.
 DR
 DR N-PSDB; AAA53723, AAA53720.
 XX
 XX Novel glycosyltransferase polypeptides and polynucleotides useful for
 PT biosynthesis of ganglioside and ganglioside mimics, as diagnostic
 PT reagents and as immunogen for producing antibodies
 PT
 XX Claim 13; Page 102-103; 120pp; English.

CC A reaction mixture for the synthesis of a sialylated oligosaccharide
 CC is useful for synthesising sialylated oligosaccharide such as
 CC ganglioside, lysoganglioside or their mimics. Glycosyltransferases
 CC are useful for chemo-enzymatic synthesis of oligosaccharides,
 CC including gangliosides and other oligosaccharides that have
 CC biological activity. The enzymes and nucleic acids that encode them
 CC are useful for studies of the pathogenesis mechanisms of organisms
 CC that synthesise ganglioside mimics, such as C. jejuni and the
 CC nucleic acids are used as probes to study expression of genes
 CC involved in ganglioside mimetic synthesis. Antibodies raised against
 CC the glycosyltransferases are also useful for analyzing the
 CC expression patterns of these genes involved in pathogenesis. The
 CC nucleic acids are also useful for designing antisense
 CC oligonucleotides for inhibiting expression of the Campylobacter
 CC enzymes that are involved in the biosynthesis of ganglioside mimics
 CC that can mask the pathogens from the host's immune system. The
 CC oligosaccharides are useful as diagnosing reagents or as therapeutics
 CC and as immunogens for producing antibodies. Bacterial
 CC glycosyltransferase can be used to catalyse the formation of
 CC oligosaccharides that are identical to the corresponding mammalian
 CC structures and are easier and less expensive to produce in large
 CC quantity, compared to the mammalian glycosyltransferase. The


```

XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 4950; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 706 AA;

```

Query Match 13.68; Score 216; DB 22; Length 706;
Best Local Similarity 23.3%; Pred. No. 3.3e-11;
Matches 79; Conservative 64; Mismatches 138; Indels 58; Gaps 10

QY 4 SIIMSVNEPLNYVRDSVESILNQTLDTEFIIVIDNPGRDGLKQFLTEYSVVDNRKITL 63
|| || || : : || || || | : : | : : : || : : || : : || : :
Db 183 slampvynveekwlrldslngvytnwelcmaddastdpnvkkilteyqqliderirvv 242

QY 64 LNEENIGLASSLNKA VKISKGEYIFRMDDDISYPSRFDKQIRFMEEN-SLDFSATLIEL 122
|| : : || : : || : : || : : || : : || : : || : : || : :
Db 243 freqnghiseatnsalalatgefvalldnddela lnafeyvkvinelpeldiysdedk 302

QY 123 IDQGKNLYVKORESNKIYLTNDIRKMLNLSLAHPTWCVKKKVPDKLMGYR-DLVVPVED 181
|| || : : : : || : : || : : || : : || : : || : : || : :
Db 303 ldmegn-----rsdpa fkpdwspdliligntnyish-lgvyrirsileeigggfkgvegsgd 355

QY 182 YDEAIRGALADFK--IGLLNKVLLOYRLNENGISOTNFKQIYYSA----- 225
|| || || || : : || : : || : : || : : || : : || : : || : :
Db 356 ydlvirftekttkerithipkvlyvrymrtpstavdgskgyaf eaglravgdalvrrgi 415

QY 226 -----ILQDFYKEK-----SYDITTKINYPQEVYIKRRYQQEELS 261
| : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 416 ngathgaanglydvyydieaklvslitptkn gykdqvrcvs sile---kttqnyei- 471

QY 262 KYFEKSPSTTIRKLXYTCLYLKFSPVLRRLINDINI 300
: : | : : | : : | : : | : : | : : | : : | : : | : :
Db 472 ----imadngstdpkmh-e-lyakkfeqqipgrffvesidi 505

RESULT 7
AAU34918
ID AAU34918 standard; Protein; 715 AA.
XX
AC AAU34918;
XX
DT 14-FEB-2002 (first entry)
XX
DE Enterococcus faecalis cellular proliferation protein #205.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Enterococcus faecalis.

XX PA (CANA) NAT RES COUNCIL CANADA.
 XX PI Gilbert M, Wakarchuk WW;
 XX DR WPI; 2000-524418/47.
 XX DR N-PSDB; AAA53720.
 XX
 PT Novel glycosyltransferase polypeptides and polynucleotides useful for
 PT biosynthesis of ganglioside and ganglioside mimics, as diagnostic
 PT reagents and as immunogen for producing antibodies
 XX
 PS Disclosure: Page 115-116; 120pp; English.
 XX
 CC A reaction mixture for the synthesis of a sialylated oligosaccharide
 CC is useful for synthesizing sialylated oligosaccharide such as
 CC ganglioside, lysoganglioside or their mimics. Glycosyltransferases
 CC are useful for chemo-enzymatic synthesis of oligosaccharides,
 CC including gangliosides and other oligosaccharides that have
 CC biological activity. The enzymes and nucleic acids that encode them
 CC are useful for studies of the pathogenesis mechanisms of organisms
 CC that synthesize ganglioside mimics, such as C. jejuni and the
 CC nucleic acids are used as probes to study expression of genes
 CC involved in ganglioside mimetic synthesis. Antibodies raised against
 CC the glycosyltransferases are also useful for analyzing the
 CC expression patterns of these genes involved in pathogenesis. The
 CC nucleic acids are also useful for designing antisense
 CC oligonucleotides for inhibiting expression of the Campylobacter
 CC enzymes that are involved in the biosynthesis of ganglioside mimics
 CC that can mask the pathogens from the host's immune system. The
 CC oligosaccharides are useful as diagnosing reagents or as therapeutics
 CC and as immunogens for producing antibodies. Bacterial
 CC glycosyltransferase can be used to catalyze the formation of
 CC oligosaccharides that are identical to the corresponding mammalian
 CC structures and are easier and less expensive to produce in large
 CC quantity, compared to the mammalian glycosyltransferase. The
 CC bacterial origin of the enzymes facilitates expression of large
 CC quantities of the enzymes used relatively inexpensive prokaryotic
 CC expression systems.
 XX
 SQ Sequence 270 AA;

Query Match 12.6%; Score 200; DB 21; Length 270;
 Best Local Similarity 26.8%; Pred. No. 2.5e-10;
 Matches 73; Conservative 52; Mismatches 119; Indels 28; Gaps 11;
 QY 4 SIIMSVNEPLNYVRDVSVESILNQTLDTEFIIVIDNPSRGDLKQFLTEYSYVDNRKIL 63
 Db 5 siilplfns-cdfisralqscinqtlkdie-iliiddkskdnsmvlefakkdprkikf 62
 QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEESLSLDFSATLIE 121
 Db 63 qneenlgtfasnlgvlnhssdfimfidsddfltpdaceiafemkkgfdllcfda-fvh 121
 QY 122 LIDQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTW--CVRK----KVFEDLMGYRD 175
 Db 122 rvtkckqyrfkqde---vfngkeflefiskrhfcswswakcfkdkilksfekikider 178
 QY 176 LVPVEDYDFAIRGALADFKIGLLNKVLLQYRLNENGISQTNKFKQYYSAILQDFYKEKS 235
 Db 179 lnygedwlfcyiyfmcfeiafvktcihyefnpg-ryenknkeiln---qnyhdkkk 233
 QY 236 YIDTTKITNYQEVVKKRYTQDEL-SKYFEL 266
 Db 234 snelik-----klskefahdefhqlflev 257
 RESULT 14
 AAB96313
 ID AAB96313 standard; Protein: 298 AA.
 XX
 AC AAB96313;

XX 29-OCT-2001 (first entry)
 XX DT Putative glycosyltransferase, involved in cell wall biogenesis #1.
 XX DE Hyperthermophilic archaeon: hyperthermophilic protein.
 XX KW Pyrococcus abyssi.
 XX OS Pyrococcus abyssi.
 XX PN FR2792651-A1.
 XX PD 27-OCT-2000.
 XX PF 21-APR-1999; 99FR-0005034.
 XX PR 21-APR-1999; 99FR-0005034.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 XX PI (IFRE-) IFREMER INST FR RECH EXPL MER.
 XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 XX PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX WPI; 2001-126236/14.
 DR New nucleotide sequences isolated from Pyrococcus abyssi encode
 PT proteins useful in industry -
 PT
 XX Claim 7; Pages 981-982; 1657pp; French.
 XX The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: This patent is in the same patent family as WO200065062, which
 CC contains additional sequences as shown in AAB99132-AAB99143,
 CC AAH75903-AAH75920 and AAG66436.
 XX
 SQ Sequence 298 AA;
 Query Match 12.5%; Score 198; DB 22; Length 298;
 Best Local Similarity 27.8%; Pred. No. 4.3e-10;
 Matches 62; Conservative 50; Mismatches 95; Indels 16; Gaps 9;
 QY 4 SIIMSVNEPLNYVRDVSVESILNQTLDTEFIIVIDNPSRGDLKQFLTEYSYVDNRKIL 63
 Db 7 sviptynr-anllrralasvlnqkfkdfe-livvddastdntpevve--siedgriryl 62
 QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEESLSLDFSATL--IE 121
 Db 63 rlknsggpiarnigikakgrfiaalldddewlphrlevqvrkfenlgkefvvygfy 122
 QY 122 LIDQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTWCVKVKVFDKLMGYRD--LVPV 179
 Db 123 yvsqdgriilgrlphkr---gdlyshllkenfispstllirrecfkk-aglfdprlss 177
 QY 180 EDYDFAIRGALADFKIGLLNKVLLQYRLNENGISQTNKFKQYI 222
 Db 178 qdwdmwlria-ryyxfdyvdeiaikyvhgkqls--fmkkyi 217
 RESULT 15
 AAY54095
 ID AAY54095 standard; Protein: 327 AA.
 XX
 AC AAY54095;
 XX
 DT 27-MAR-2000 (first entry)
 XX

DE Enzyme EPSJ involved in exopolysaccharide biosynthesis.

XX Exopolysaccharide; EPS; ESP enzyme; EPSA; EPSB; EPSB; EPSD; EPSF; EPSG; EPSH; EPSI; EPSJ; EPSK; Lactococcus delbrueckii bulgaricus; KW activated D-galactose pyranose; saccharide; beta-galactosyltransferase; KW transcription attenuator; glucosyl-phosphotransferase; KW alpha-galactosyltransferase; glucosyltransferase; EPSL; EPSM; KW alpha-galactosyltransferase; EPS polymerase; glycosyltransferase; EPSN; KW phosphofuranose; transporter; food; fermented milk product; yoghurt; KW cheese; flavour stability; organoleptic property.

OS *Lactobacillus delbrueckii* bulgaricus.

AA PN W09962316-A2.

XX
PD
09-DEC-1999XX
PF 22-APR-1999: 99WO-EP02841.XX
PR 22-APR-1998: 98EP-0201310.

PR	22-APR-1998;	98EP-0201310;
PR	22-APR-1998;	98EP-0201311.

PR 22-APR-1998; 98EP-0201312.

PA (NEST.) SOC PROD NESTLE SA.

PI Stingeles F, Germond JE, Lamothe G:

AA
DR WPI: 2000-097267/08.

DR N-PSDB; AAZ45260;

New recombinant enzymes for synthesis of exopolysaccharides, particularly in lactic acid bacteria, for improving properties of fermented milk products -

PS Example 18; Page 156-157; 162pp; French.

AA54086-99 represent enzymes involved in the biosynthesis of exopolysaccharides (EPS). These enzymes are designated EPSA-EPSN, and are encoded by open reading frames epsA-epsN. The enzymes are isolated from *Lactobacillus delbrueckii* bulgaricus. The proteins are used in a method for the synthesis of EPS, which includes at least one step of forming a bond (alpha or beta-isomer) between C-1 (carrying the reducing aldehyde function, of an activated D-galactose pyranose), and a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis of EPS occurs with, in each step, addition of a new sugar unit, through its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar unit, present at the end of a chain of sugar residues bonded to the primer. EPSA is attenuator of transcription which influences regulation. EPSB determines the length of the saccharide chain; EPSC is involved in the control of the molecular weight and/or the length of the polysaccharide; EPD, EPL and EPSM are involved in synthesis of EPS; EPSE is a galactosyl- or glucosyl-phospho-transferase which catalyses the transfer of the first saccharide on the primer; EPSF and EPSG are alpha-glucosyltransferases; EPSH and EPSHJ are beta-glucosyltransferases; EPSG is a glucosyltransferase; EPSK is responsible for the polymerisation of the repetitive units; and EPSN is responsible for export of the EPS. The EPS enzyme are used to improve properties of foods, particularly fermented milk products such as yoghurt and cheese, e.g. their organoleptic properties and flavour stability.

XX	Sequence	327 AA:
SO		

Query Match	12.2%	Score 193.5;	DB 21;	Length 327;
Best Local Similarity	25.0%;	Pred. No. 1.3e-09;		
Matches 83;	Conservative	55;	Mismatches 121;	
			Indels	73;
			Gaps	15;

QY 1 MNYSIIMSVYNEPLNYVRDSVESILNOTLTDFEEFIIVIDNPSRGDLKOFLTEYSVVDNRI 60

[illegible]

QY 61 KILLNEENIGLASSINKAVKISKGEYIFRMDADDISYPFRDQKIFPMEEENSLEFSA TL 120

Db	59	kv-vhkngqglsarnagmkvatgeyisfidssdytasmyehvufsimkcecadivvgr	117
Qy	121	ELIDQGNLYVKQRESNKIYLTNDIRK--MLLNRSILAHPTWCVKKKVDFDKLGMGYBDLP	178
Db	118	cyvypngs--kkirekqvnevmdgpkataimnts-----llgyfdaaa	159
Qy	179	VEDYDFAIRGALADFKLGLLNKVL--QYRLNENGISOTNKF---KQYIYSATLODYPKEK	234
Db	160	wdkv-----ykrslfndvsyspegklisedwytykafkanrlvdydstpmvyygr	209
Qy	235	--SYIDITKITNTYFQEVV-----TKKR--YTOOELSKYF-----ELKSTPSIRI	274
Db	210	ggsitststvnjdanyasrevdyfkkkqpeytaeanfayfsrfgvidnlrtvqptvdk	269
Qy	275	RKLYCYLYLYFKSPLVRRLIINDI--NILVLK	304
Db	270	qki-----rkirndmkanikalk	287

Search completed: July 31, 2002, 18:56:02
Job time: 343 sec